ANALYSIS OF HIGHER ORDER MARKOV AND SEMI-MARKOV CHAINS

By

HATEM AHMED ELAYAT

Bachelor of Science
Ain Shams University
Cairo, Egypt
1965

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1970

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
July, 1973

ANALYSIS OF HIGHER ORDER MARKOV AND SEMI-MARKOV CHAINS

Thesis Approved:

James E. Shamblin
Thesis Adviser

Hamed K. Eldin

R. Leroy Folks

G. T. Stevens, Jr.

N. N. Durham
Dean of the Graduate College

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my gratitude to those individuals without whose help and encouragement the attainment of this level of education would not have been possible. Primary among those are the members of my committee: Dr. Hamed K. Eldin, Dr. James E. Shamblin, Dr. J. Leroy Folks, and Dr. G. T. Stevens.

I wish to express my sincere gratitude to Dr. Hamed K. Eldin for serving as chairman of my advisory committee and for his inspiration and advice during my graduate study. I want to express my personal gratitude to Dr. James E. Shamblin, my thesis adviser, who suggested this thesis topic and provided help and encouragement during its development. I am deeply indebted to Dr. J. Leroy Folks for his continual source of helpful suggestions. Also, I would like to thank Dr. Stevens for his guidance and encouragement.

Mrs. Margaret Estes has done an excellent job of typing this paper, and I thank her for this.

Above all, I would like to express my deep gratitude to my parents, for their continued encouragement, understanding, and numerous sacrifices during the attainment of this degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Markov Processes were first defined and studied by the Russian mathematician A. A. Markov (1856-1922). Since then the theory has been extensively developed, and has found applications in many fields such as Operations Research, Economic studies, Nuclear Physics, Educational Psychology, and Genetics.

First Order Markov Chains

The problem of analyzing first order Markov chains has received a considerable amount of attention in the literature. Research works have been contributed by Kemeny and Snell (26), Feller (13), Howard (22), Chung (9), and many others. A short description of first order, discrete parameter Markov chains will follow.

Consider a system which is observed at a discrete set of times. Let the successive observations be denoted by $X_0, X_1, \ldots, X_n, \ldots$. It is assumed that X_n is a random variable. The value of X_n represents the state at time n of this system. The sequence $\{X_n\}$ is called a first order discrete parameter Markov chain if the following conditions are satisfied:

1) The number of states in the system is finite.

2) Each random variable X_n is discrete.

3) For any integer $m > 2$ and any set of m points $n_1 < n_2 < \ldots < n_m$

the conditional distribution of X_{n_m}, for given values of $X_{n_1}, \ldots, X_{n_{m-1}}$, depends only on $X_{n_{m-1}}$, the most recent known value; in particular, for any real numbers $x_0, x_1, \ldots, x_m$, one obtains

$$P\left[X_m = x_m \mid X_0 = x_0, \ldots, X_{m-1} = x_{m-1}\right] = P\left[X_m = x_m \mid X_{m-1} = x_{m-1}\right] \qquad (1.1)$$

this is known as the Markov property.

This means that knowledge of $X_0, X_1, \ldots, X_{m-1}$ gives no more information for predicting the value of X_m than does knowledge of X_{m-1} alone. In other words, the system has no "memory" that would allow it to use information about its behavior, before a known state was reached, to modify the probabilities for the next stage. In essence this requires that it be possible to deduce the future development of the process from knowledge of its present state. Information about the history of the process has no predictive value.

This is a severe and sometimes unrealistic restriction, as stated by Bartholomew (3). Howard (22) states that there are few physical systems that one could expect to be so memoryless in a strict sense. Cox (11) adds that this is a very strong restriction on the process. Kuehn (30), in application to Consumer Brand switching, analyzed a sequence of five purchases to determine the influence of the consumer's previous purchases on his next purchase. He concluded that the most recent purchase of the consumer is not the only one influencing his brand choice, but rather previous purchases had an effect. He accordingly questions the validity of the applicability of first order Markov chains to such a case.

Systems, in which information about their history enables one to deduce their future development, are better described or modeled as higher order Markov chains.

Higher Order Markov Chains

Higher order Markov chains are those Markov chains whose future outcomes depend upon one or more immediately preceding states. For example, in the case of the first order Markov chain the next outcome depends only upon the present state, in the second order Markov chain the future outcome depends upon the present state and the state immediately preceding the present state. In the kth order chains the future outcome will depend upon the present state and the (k-1) states immediately preceding the present state. Considering a sequence of trials, the outcome of each trial depends only on the outcomes of the k directly preceding trials. The sequence of random variables $\{X_n\}$ mentioned previously forms a Markov chain of order k if, given a fixed k, for all n and for all possible values of the variables, it is true that

$$P\left[X_m = x_m \mid X_0 = x_0, \ldots, X_{m-1} = x_{m-1}\right] = P\left[X_m = x_m \mid X_{m-k} = x_{m-k}, \ldots, X_{m-1} = x_{m-1}\right]. \tag{1.2}$$

Little research has been done in the area of higher order Markov chains. Cox (11) and Howard (22) have considered the problem of analyzing higher order Markov chains. They proposed a method by which one would reduce the process to a Markov chain by appropriately redefining the state space, with this redefinition, considerable increase in computations would arise. Ganesan (16) considered analyzing higher order Markov chains using n-dimensional matrices. Anderson and Goodman (1)

considered the problem of estimating the transition probabilities of higher order Markov chains, which was also considered by Telser (40).

Statement of the Problem

The purpose of this research is to develop appropriate methods of analysis for higher order Markov and Semi-Markov chains, thus allowing them to be applied to various problems in different areas.

Higher order Markov chains will be analyzed using n-dimensional matrices. Chapter II introduces the mathematics of n-dimensional matrices. The different operations of addition, multiplication and the inverse of n-dimensional matrices are defined. An algorithm for finding the inverse of n-dimensional matrices is developed.

Chapter III is devoted to deriving the Chapman Kolmogorov equation. An algorithm to compute the steady state probabilities is developed.

Chapter IV investigates the absorption characteristics of higher order Markov chains. The following quantities which are of major interest in the study of absorbing Markov chains are developed:

1) The expected number of steps before the process is absorbed.
2) The expected number of times the process is in a given non-absorbing state.
3) The probability of absorption by any given absorbing state.

Chapter V deals with the maximum likelihood method of estimation of the transition probability matrix. Also, the problem of testing the hypothesis that the chain is of a given order is discussed.

Chapter VI introduces higher order Semi-Markov chains and their analysis. Both the discrete-time Semi-Markov processes and the

continuous-time Semi-Markov processes are considered. Emphasis is placed on finding the holding and waiting time statistics.

CHAPTER II

ANALYSIS OF HIGHER ORDER MARKOV CHAINS USING n-DIMENSIONAL MATRICES

The primary purpose of this chapter is to introduce the concept of n-dimensional matrices, how they are used to analyze higher order Markov chains and their computational advantages.

Mathematics of n-Dimensional Matrices

The theory of application of n-dimensional matrices has received little attention in the literature. Sikorski (39), Kron (29), and Ganesan (16) have considered some of the properties of n-dimensional matrices. This section is concerned with the theory of n-dimensional matrices.

Denote an n-dimensional matrix A by

$$A = \left[a_{i_1,\ldots,i_n}\right]_{(m,m,\ldots,m)} \quad \text{, where } i_1,i_2,\ldots,i_n = 1,2,\ldots,m \qquad (2.1)$$

(n times)

then the numbers $a_{i_1,i_2,\ldots,i_n}$ are said to be the coordinates (or elements) of the matrix A; more precisely, the element $a_{i_1,\ldots,i_n}$ is said to be the $(i_1,\ldots,i_n)$-th coordinate of the matrix A.

For example a three-dimensional matrix A will be denoted by

$$A = \left[a_{i_1 i_2 i_3}\right]_{(m,m,m)}$$

and this is represented in Figure 1.

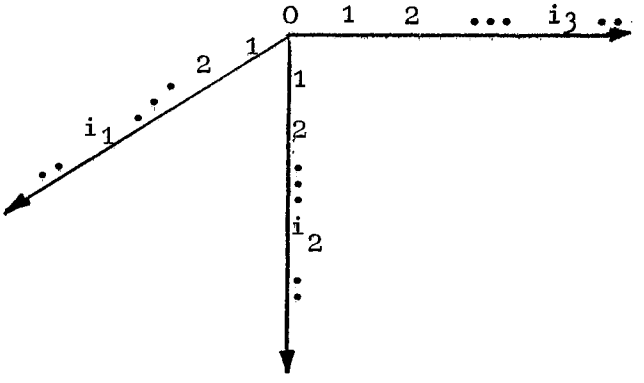

Figure 1. A Representation of a Three-Dimensional Matrix Subscripted by (i_1, i_2, i_3)

It is not possible to display a matrix in the full n-dimensional space. A two-dimensional array identified by subscripts will be used instead. So the matrix of the previous example for the case $m = 2$, will be shown as follows

$$\underset{(2 \times 2 \times 2)}{A} = \begin{bmatrix} a_{111} & a_{112} \\ a_{121} & a_{122} \\ a_{211} & a_{212} \\ a_{221} & a_{222} \end{bmatrix} .$$

We shall now define some algebraic operations on n-dimensional matrices.

Addition of Matrices

If

$$A = \left[a_{i_1,\ldots,i_n}\right]_{(m,\ldots,m)} \quad \text{and} \quad B = \left[b_{i_1,\ldots,i_n}\right]_{(m,\ldots,m)}, \tag{2.2}$$

we define the sum A + B to be the matrix

$$A + B = \left[a_{i_1,\ldots,i_n} + b_{i_1,\ldots,i_n}\right]_{(m,\ldots,m)}.$$

That is, the sum of two matrices of the same order is found by adding the corresponding elements thereof.

Commutative and Associative Laws of Addition

The addition of matrices is both commutative and associative, that is, if A, B, and C are conformable for addition,

$$A + B = \left[a_{i_1,\ldots,i_n} + b_{i_1,\ldots,i_n}\right] = \left[b_{i_1,\ldots,i_n} + a_{i_1,\ldots,i_n}\right] = B + A$$

and

$$\begin{aligned} A + (B+C) &= \left[a_{i_1,\ldots,i_n} + (b_{i_1,\ldots,i_n} + c_{i_1,\ldots,i_n})\right] \\ &= \left[(a_{i_1,\ldots,i_n} + b_{i_1,\ldots,i_n}) + c_{i_1,\ldots,i_n}\right] = (A+B) + C . \end{aligned}$$

Subtraction of Matrices

In a similar manner we define the difference A-B of the matrices defined in (2.2)

$$A - B = \left[a_{i_1,\ldots,i_n} - b_{i_1,\ldots,i_n}\right]_{(m,\ldots,m)}.$$

Multiplication by a Scalar

Given a scalar c, and matrix A, as defined previously, then

$$cA = \left[ca_{i_1,\ldots,i_n}\right]_{(m,\ldots,m)} .$$

The Multiplication of Matrices

If

$$A = \left[a_{i_1,\ldots,i_n}\right]_{(m,\ldots,m)}, \quad B = \left[b_{i_1,\ldots,i_n}\right]_{(m,\ldots,m)},$$

and

$$C = \left[c_{i_1,\ldots,i_n}\right]_{(m,\ldots,m)} \tag{2.3}$$

we define the product C = AB, where

$$c_{i_1 i_2,\ldots,i_{n-1} i_n} = \sum_{k=1}^{m} a_{i_1 i_2,\ldots,i_{n-1} k}\, b_{i_2 i_3,\ldots,i_{n-1} k i_n} . \tag{2.4}$$

The Properties of Matrix Multiplication

For the matrices defined in (2.3), we summarize the three fundamental properties of matrix multiplication:

(a) The commutative law does <u>not</u> hold true generally: $AB \neq BA$.

(b) The associative law does <u>not</u> hold: $(AB)C \neq A(BC)$.

(c) Matrix multiplication is distributive with respect to addition. This means $A(B+C) = AB + AC$. Also if we define the matrices D, E, and F in a similar manner, then $(D+E)F = DF + EF$.

The following equation follows from the definition of multiplication given by (2.4) and from the above properties:

$$A^k = AA^{k-1} \neq A^{k-1}A \quad \text{for} \quad k = 2, 3, \ldots \ . \tag{2.5}$$

The Identity Matrix

We define the identity matrix I such that for any matrix P

$$PI = P \tag{2.6}$$

but we note that $IP \neq P$. An example of a three-dimensional identity matrix with $m = 2$ is given by

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} .$$

Matrix Inversion

Given the two n-dimensional matrices A and B where

$$A = \left[a_{i_1,\ldots,i_n}\right]_{(m,\ldots,m)} \quad \text{and} \quad B = \left[b_{i_1,\ldots,i_n}\right]_{(m,\ldots,m)}$$

we define the matrix B to be the inverse of A if

$$BA = I \ . \tag{2.7}$$

Due to the properties of matrix multiplication mentioned earlier $AB \neq I$.

Before the algorithm for inverting the n-dimensional matrix is presented, the partitioning will be discussed.

Partition the n-dimensional matrix A into $m^{(n-2)}$ two-dimensional matrices as follows:

$$A = \begin{bmatrix} A_1^* \\ A_2^* \\ \vdots \\ A_s^* \end{bmatrix}, \quad \text{where } s = m^{(n-2)} \text{ and } (n > 2) \ . \tag{2.8}$$

A_j^* is an m by m two-dimensional matrix, (j = 1,2,...,s). For example, if A is a three-dimensional matrix, and m = 3, n = 3, then we can partition this matrix to $3^{2-1} = 3$ two-dimensional matrices as follows:

$$A = \begin{bmatrix} a_{111} & a_{112} & a_{113} \\ a_{121} & a_{122} & a_{123} \\ a_{131} & a_{132} & a_{133} \\ a_{211} & a_{212} & a_{213} \\ a_{221} & a_{222} & a_{223} \\ a_{231} & a_{232} & a_{233} \\ a_{311} & a_{312} & a_{313} \\ a_{321} & a_{322} & a_{323} \\ a_{331} & a_{332} & a_{333} \end{bmatrix}$$

$$A_1^* = \begin{bmatrix} a_{111} & a_{112} & a_{113} \\ a_{121} & a_{122} & a_{123} \\ a_{131} & a_{132} & a_{133} \end{bmatrix}, \qquad A_2^* = \begin{bmatrix} a_{211} & a_{212} & a_{213} \\ a_{221} & a_{222} & a_{223} \\ a_{231} & a_{232} & a_{233} \end{bmatrix},$$

$$A_3^* = \begin{bmatrix} a_{311} & a_{312} & a_{313} \\ a_{321} & a_{322} & a_{323} \\ a_{331} & a_{332} & a_{333} \end{bmatrix} .$$

Algorithm:

1 - Partition the matrix A into $m^{(n-2)}$ two-dimensional matrices of equal size, $A_1^*, A_2^*, \ldots, A_s^*$, where $s = m^{(n-2)}$.

2 - Invert A_j^* (for $j = 1,2,\ldots,s$).

3 - Generate a sequence of numbers $123\ldots m12\ldots m12\ldots m, m^{(n-3)}$ times; for example, if $n = 5$, $m = 3$, then we have $m^{(n-3)} = 3^{5-3} = 9$ and we get the following sequence

123	123	123	123	123	123	123	123	123
1	2	3	.	.	.	.	.	9

4 - The matrix $B = A^{-1}$ is given by

$$B = \begin{bmatrix} B^* \\ B^* \\ \vdots \\ B^* \end{bmatrix}$$, we have mB^* submatrices

where B^* is an $m^{(n-2)}$ by m two-dimensional matrix.

The elements of B^* are obtained as follows:

The 1st row of B^* will be the 1st row of A_1^{*-1}

The 2nd row of B^* will be the 2nd row of A_2^{*-1}

$\vdots$ $\vdots$

The mth row of B^* will be the mth row of A_m^{*-1}

The (m+1)th row of B^* will be the 1st row of A_{m+1}^{*-1}

The (m+2)th row of B^* will be the 2nd row of A_{m+2}^{*-1}

and so on according to the generated sequence.

For example if $m = 3$, $n = 4$ we have $m^{n-3} = 3^{4-3} = 3$ the generated sequence is 123123123

$\therefore$ The 1st row of B^* will be the 1st row of A_1^{*-1}

The 2nd row of B^* will be the 2nd row of A_2^{*-1}

The 3rd row of B^* will be the 3rd row of A_3^{*-1}

The 4th row of B^* will be the 1st row of A_4^{*-1}

The 5th row of B^* will be the 2nd row of A_5^{*-1}

The 6th row of B^* will be the 3rd row of A_6^{*-1}

The 7th row of B^* will be the 1st row of A_7^{*-1}

The 8th row of B^* will be the 2nd row of A_8^{*-1}

The 9th row of B^* will be the 3rd row of A_9^{*-1} .

We note that the n-dimensional matrix A has an inverse if and only if A_j^* (for $j = 1, \ldots, s$) are non singular.

Example 2-1

Find the inverse of the following three-dimensional matrix A.

$$A = \begin{bmatrix} 4 & 1 & 2 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \\ 5 & 3 & 2 \\ 2 & 4 & 1 \\ 2 & 1 & 2 \\ 3 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

Steps

1 - Partition the matrix A into $m^{(n-2)}$ two-dimensional matrices.

$$m^{(n-2)} = 3^{3-2} = 3 \text{ - matrices}$$

$$A_1^* = \begin{bmatrix} 4 & 1 & 2 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix} \quad A_2^* = \begin{bmatrix} 5 & 3 & 2 \\ 2 & 4 & 1 \\ 2 & 1 & 2 \end{bmatrix} \quad A_3^* = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} .$$

2 - Invert A_j^* (for $j = 1,2,3$)

$$A_1^{*-1} = \begin{bmatrix} 1 & 0 & -1 \\ -\frac{1}{5} & \frac{2}{5} & 0 \\ -\frac{7}{5} & -\frac{1}{5} & 2 \end{bmatrix} \quad A_2^{*-1} = \begin{bmatrix} \frac{7}{17} & -\frac{4}{17} & \frac{5}{17} \\ -\frac{2}{17} & \frac{6}{17} & -\frac{1}{17} \\ -\frac{6}{17} & -\frac{1}{17} & \frac{14}{17} \end{bmatrix} \quad A_3^{*-1} = \begin{bmatrix} \frac{3}{7} & \frac{1}{7} & -\frac{1}{7} \\ -\frac{1}{7} & \frac{5}{7} & -\frac{2}{7} \\ -\frac{1}{7} & -\frac{2}{7} & \frac{5}{7} \end{bmatrix}.$$

3 - Generate a sequence of numbers

$1\ 2\ ..\ m\ 1\ 2\ ..\ m$ $\quad m^{(n-3)}$ times

$m^{(n-3)} = 3^{3-3} = 1$ time

the sequence is 1 2 3.

4 - The matrix $B=A^{-1}$ is given by

$$B = \begin{bmatrix} B^* \\ B^* \\ B^* \end{bmatrix}$$ we have mB^* ; i.e., $3B^*$

where B^* is an $m^{(n-2)}$ by m two-dimensional matrix; i.e., 3 by 3.

The 1st row of B^* will be the 1st row of A_1^{*-1} $[1 \quad 0 \quad -1]$.

The 2nd row of B^* will be the 2nd row of A_2^{*-1} $\left[-\frac{2}{17} \quad \frac{6}{17} \quad -\frac{1}{17}\right]$.

The 3rd row of B^* will be the 3rd row of A_3^{*-1} $\left[-\frac{1}{7} \quad -\frac{2}{7} \quad \frac{5}{7}\right]$.

$\therefore$

$$\underset{(3 \times 3)}{B^*} = \begin{bmatrix} 1 & 0 & -1 \\ -\frac{2}{17} & \frac{6}{17} & -\frac{1}{17} \\ -\frac{1}{7} & -\frac{2}{7} & \frac{5}{7} \end{bmatrix}$$

$$\therefore \quad A^{-1} = B = \begin{bmatrix} 1 & 0 & -1 \\ -\frac{2}{17} & \frac{6}{17} & -\frac{1}{17} \\ -\frac{1}{7} & -\frac{2}{7} & \frac{5}{7} \\ 1 & 0 & -1 \\ -\frac{2}{17} & \frac{6}{17} & -\frac{1}{17} \\ -\frac{1}{7} & -\frac{2}{7} & \frac{5}{7} \\ 1 & 0 & -1 \\ -\frac{2}{17} & \frac{6}{17} & -\frac{1}{17} \\ -\frac{1}{7} & -\frac{2}{7} & \frac{5}{7} \end{bmatrix}.$$

Check

$$A^{-1}A = I$$

$$\begin{bmatrix} 1 & 0 & -1 \\ -\frac{2}{17} & \frac{6}{17} & -\frac{1}{17} \\ -\frac{1}{7} & -\frac{1}{7} & \frac{5}{7} \\ 1 & 0 & -1 \\ -\frac{2}{17} & \frac{6}{17} & -\frac{1}{17} \\ -\frac{1}{7} & -\frac{2}{7} & \frac{5}{7} \\ 1 & 0 & -1 \\ -\frac{2}{17} & \frac{6}{17} & -\frac{1}{17} \\ -\frac{1}{7} & -\frac{2}{7} & \frac{5}{7} \end{bmatrix} \times \begin{bmatrix} 4 & 1 & 2 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \\ 5 & 3 & 2 \\ 2 & 4 & 1 \\ 2 & 1 & 2 \\ 3 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Application of n-Dimensional Matrices to Higher Order Markov Chains

We shall denote $p_{i_1 i_2 \cdots i_k i_{k+1}}$ to be the conditional probability of the process being in state i_{k+1} at the nth step given that it was in the states $i_1, i_2, \ldots, i_k$ at $t = 0,1,2,\ldots,n-1$, where $(i_1, i_2, \ldots, i_{k+1} = 1,2,\ldots,m)$ and $(k = 1,2,\ldots)$ being the order of the chain.

Let $(X_t;\ t = 0,1,\ldots)$ be a second order Markov chain with the state space S having only two discrete points (1,2), then its three-dimensional transition probability matrix P in two dimensions will be given below:

$t = 0$	$t = 1$	$t = 2$	
		1	1
1	1	P_{111}	P_{112}
	2	P_{121}	P_{122}
2	1	P_{211}	P_{212}
	2	P_{221}	P_{222}

Each row in P is a probability vector describing the process exhaustively for the given present state and the state immediately preceding the present state. The elements of this transition matrix $P = \left[p_{i_1 i_2 i_3}\right]_{(2,2,2)}$ must satisfy the following:

$$0 \leq p_{i_1 i_2 i_3} \leq 1$$

$$\sum_{i_3=1}^{2} p_{i_1 i_2 i_3} = 1 \quad \text{for} \quad (i_1, i_2 = 1,2)\ . \tag{2.9}$$

This matrix P is shown in Figure 2 in three dimensions. In Figure 2, there are 2 X 2 matrices, one for every possible state at $t = 0$. If the

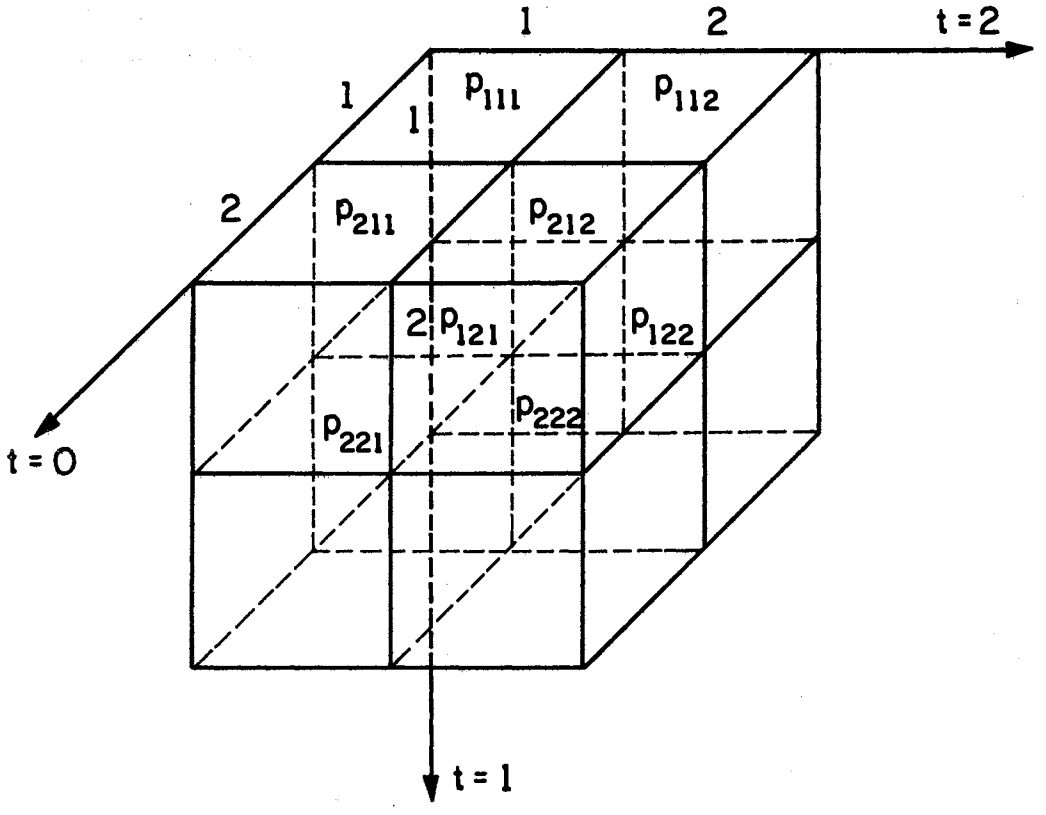

Figure 2. The Transition Probability Matrix P of a Second Order Markov Chain with the State Space S = (1,2)

process does not depend upon the outcome at time $t = 0$, then these 2×2 matrices reduce to a single 2×2 matrix independent of the outcome at $t = 0$ and it is a first order Markov chain.

For the general kth order Markov chains with m states, the transition probability matrix is given by

$$P = \left[p_{i_1 \ldots i_{k+1}} \right]_{(m,\ldots,m)}, \tag{2.10}$$

where

$$0 \leq p_{i_1 \ldots i_{k+1}} \leq 1$$

$$\sum_{i_{k+1}=1}^{m} p_{i_1 \ldots i_{k+1}} = 1 \quad \text{for} \quad (i_1, \ldots, i_k = 1,2,\ldots,m).$$

This matrix will have m^{k+1} elements. As in the case of first order Markov chains, the elements of this matrix are also called "one step transition probabilities" since they describe the conditional probability of being in a particular state in the nth step, given the states at $t = 0,1,\ldots,n-1$.

Computational Advantages

Howard (22) and Cox (11) both proposed a method for analyzing higher order Markov chains. In this section we shall present their method of analysis and compare it to the proposed method using n-dimensional matrices and discuss the computational advantages the proposed method offers.

Cox (11) states that processes may be encountered in which the dependence goes back more than one time unit and in such cases the process can be reduced to a Markov chain by appropriately redefining

the state space or changing the state structure. An example may best illustrate the procedure.

Example 2-2

Consider a sequence of dependent Bernoulli trials in which the probability of success, denoted by 1, or failure, denoted by 0, at any given trial depends on the outcome of the two preceding trials. Thus we have a two state second order Markov chain. We redefine the state space according to the outcome of two successive trials:

00	state 0
01	state 1
10	state 2
11	state 3 .

Thus if trials (n-1) and n give rise, for example to 1 and 0, then we say that the process is in state 2 at time n. It is easy to see that the process is now a Markov chain with four states.

For the general second order Markov chains with m states, we can define a new process (first order) with m^2 states, each state in the new process would correspond to a pair of successive states in the old process. With this redefinition a considerable increase in computational complexity would arise. The transition probability matrix would have m^4 elements.

For the general kth order Markov chains with m states, the new defined process (1st order) will have m^k states, and the transition probability matrix will have m^{2k} elements.

As mentioned earlier, when using n-dimensional matrices for analyzing kth order Markov chains with m states, the transition probability matrix will have m^{k+1} elements only. This results in a

reduction in the number of elements of the transition probability matrix, and is given by

$$m^{2k} - m^{k+1} = m^{k+1}(m^{k-1} - 1) \ . \qquad (2.11)$$

Figure 3 presents the plot of such a reduction for second order Markov chains.

As a result of the reduction in the number of elements of the transition probability matrix, considerable amount of computations will be reduced.

For the problem of matrix multiplication, we know that for multiplying two square matrices each of size n by n, then the process of multiplying these two matrices requires

$$n^3 \text{ multiplications, } n^3 \text{ additions} \ . \qquad (2.12)$$

Not counting additions, the reduction in the number of multiplications is given by

$$m^{3k} - m^3 \cdot m^{k-1} = m^k(m^{2k} - m^2) \ . \qquad (2.13)$$

For the problem of matrix inversion, we know that for a square matrix of size n, the inversion of this matrix by the Gauss process requires, as indicated by Fox (14):

$$n \text{ reciprocals, } n^3 - 1 \text{ multiplications, } n^3 - 2n^2 + n \text{ additions.} \qquad (2.14)$$

Considering only the number of multiplications, the reduction in computation is approximately given by (2.13). Figure 4 presents a plot of this computational reduction for second order Markov chains.

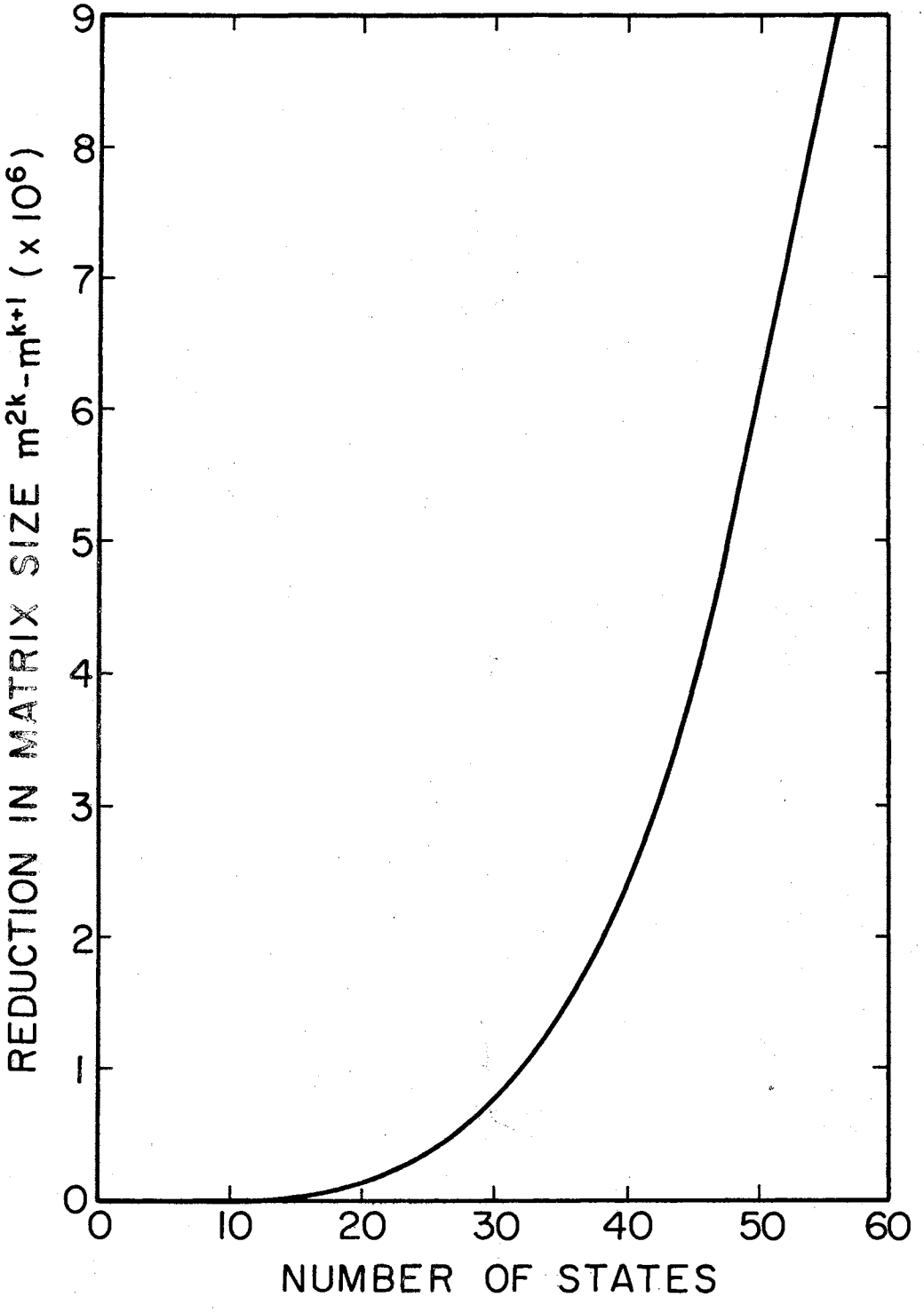

Figure 3. Element Reduction for Second Order Markov Chains as a Function of the Number of States

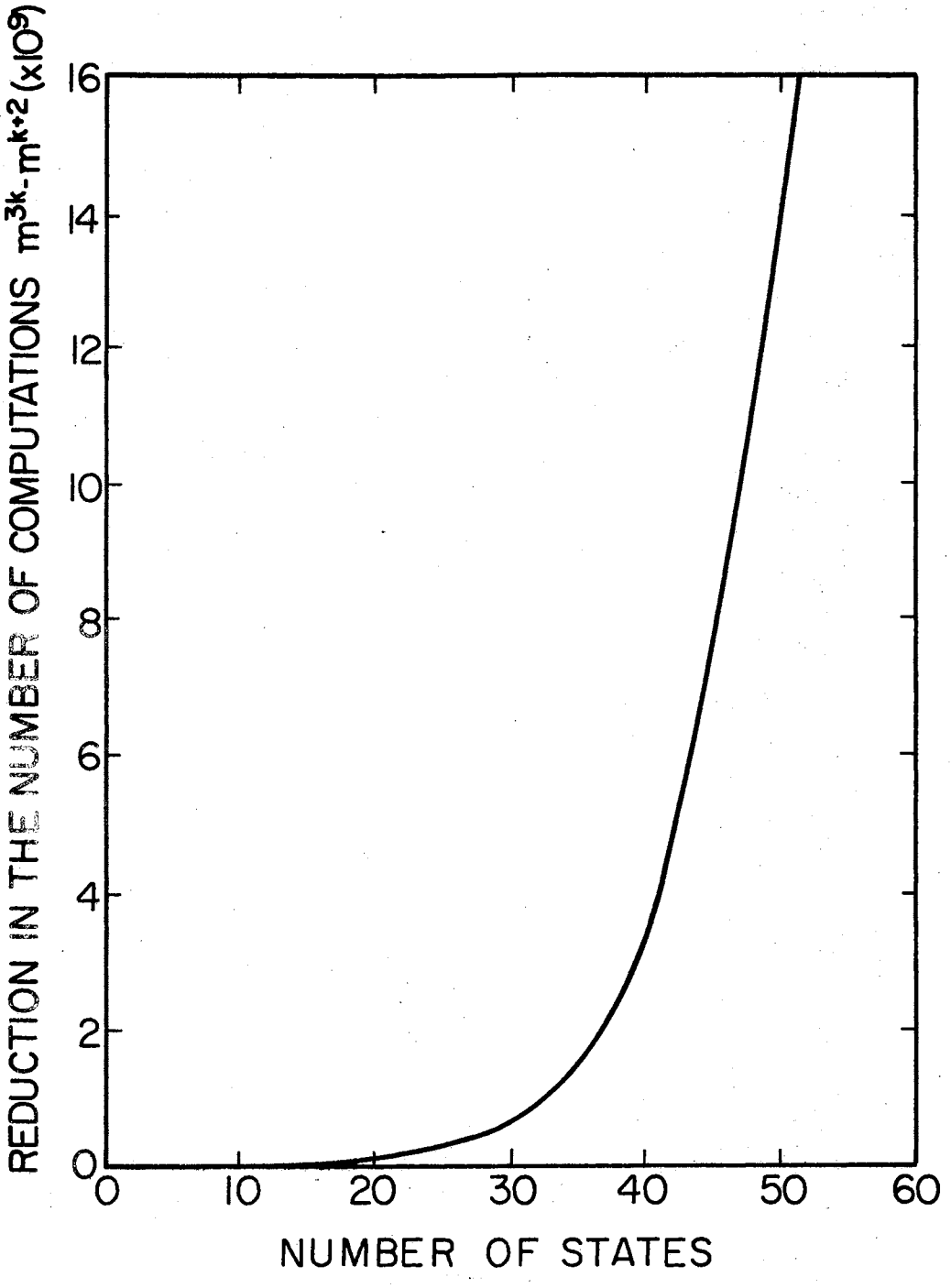

Figure 4. Computational Reduction for Second Order Markov Chains as a Function of the Number of States

CHAPTER III

THE STEADY STATE CONDITIONS FOR HIGHER ORDER ERGODIC MARKOV CHAINS

This chapter will be divided into two sections. First the Chapman-Kolmogorov equations for higher order Markov chains will be derived. Then, a development of the steady state Algorithm for such chains will be presented.

Chapman-Kolmogorov Equations

We will adopt the following notation:

P - Transition probability matrix, or matrix of one step probabilities.

$p_{i_1 \cdots i_{k+1}}$ - An element of the matrix of one-step transition probabilities. It is the conditional probability of the process being in state i_{k+1} at the nth step given that it was in the states $i_1, i_2, \ldots, i_k$ at $t=0,1,2,\ldots,n-1$, where $(i_1, i_2, \ldots, i_{k+1}=1,2,\ldots,m)$ and $(k=1,2,\ldots)$ being the order of the chain, and m the number of states.

$p^s_{i_1 \cdots i_{k+1}}$ - Probability of the process being in the state i_{k+1} at the $(n-1+s)$th step given that it was in the states $i_1, i_2, \ldots, i_k$ at $t=0,1,2,\ldots,n-1$, respectively. These are also known as the s-step transition probabilities.

The derivation of the Chapman-Kolmogorov equation will be given for the second order chains. For higher order chains, the proof follow in a similar manner.

The Chapman-Kolmogorov equation for second order chain is given by

$$p^{(s)}_{i_1 i_2 i_3} = \sum_{r=1}^{m} p_{i_1 i_2 r}\, p^{(s-1)}_{i_2 r i_3} \tag{3.1}$$

or in other words

$$p^{(1+s)}_{i_1 i_2 i_3} = \sum_{r=1}^{m} p_{i_1 i_2 r}\, p^{(s)}_{i_2 r i_3} \quad . \tag{3.2}$$

Proof: The proof is by induction

1 - For $s = 1$ Equation (3.2) becomes

$$p^{(1+1)}_{i_1 i_2 i_3} = \sum_{r=1}^{m} p_{i_1 i_2 r}\, p_{i_2 r i_3}$$

this is true from the law of total probability.

2 - Assume Equation (3.2) is true for $s = h$, where h is any positive integer

$$\therefore \quad p^{(1+h)}_{i_1 i_2 i_3} = \sum_{r=1}^{m} p_{i_1 i_2 r}\, p^{(h)}_{i_2 r i_3} \quad .$$

3 - For $s = h + 1$, Equation (3.2) becomes

$$p^{(1+(h+1))}_{i_1 i_2 i_3} = \sum_{r=1}^{m} p_{i_1 i_2 r}\, p^{(h+1)}_{i_2 r i_3}$$

but this is true from the assumption in step 2 and from the law of total probability.

4 - Since the statement is true for $s = 1$ (from step 1) it must (from step 3) be true for $s = 1 + 1 = 2$ and from this for

$s = 2 + 1 = 3$, etc., and so must be true for all positive integers.

$\therefore\ p^{(1+s)}_{i_1 i_2 i_3} = \sum_{r=1}^{m} p_{i_1 i_2 r}\, p^{(s)}_{i_2 r i_3}$ is true for all values of s,

or in other words

$$p^{(s)}_{i_1 i_2 i_3} = \sum_{r=1}^{m} p_{i_1 i_2 r}\, p^{(s-1)}_{i_2 r i_3} .$$

Equation (3.1) is a statement of the fact that in passing from state i_2 to state i_3 in s steps given that the process was in state i_1 the time before, the first step takes the process from state i_2 to some intermediate state r and the remaining (s-1) steps then takes the process from r to i_3.

Similarly, it can be shown that the Chapman-Kolmogorov equation for the kth order chain is given by

$$p^{(s)}_{i_1 i_2 \cdots i_k i_{k+1}} = \sum_{r=1}^{m} p_{i_1 \cdots i_k r}\, p^{(s-1)}_{i_2 i_3 \cdots i_k r i_{k+1}} . \qquad (3.3)$$

From (3.3), it follows that

$$P^{(s)} = P\, P^{(s-1)} , \qquad (3.4)$$

where

$$P^{(s)} = \left[p^{(s)}_{i_1 \cdots i_{k+1}} \right]_{(m,\ldots,m)} . \qquad (3.5)$$

This means that

$$P^{(1)} = P$$

$$P^{(2)} = P\, P = P^2$$

$$P^{(3)} = P\, P^{(2)} = P(P\, P) = P^3$$

$$P^{(4)} = P\,P^{(3)} = P(P(P\,P)) = P^4$$

$$\vdots \qquad\qquad \vdots$$

$$P^{(s)} = P\,P^{(s-1)} = P(P(P\ldots(P(P\,P))\ldots)) = P^s \,. \qquad (3.6)$$

As mentioned in Chapter II, the associative law for multiplication does not hold for n-dimensional matrices, then we have to notice the order of multiplication as indicated by brackets in (3.6) and we note also that

$$P^{(a+b)} \neq P^{(a)}\,P^{(b)} \text{ where a,b are positive integers.}$$

The Steady State Algorithm

Definition 3-1: An ergodic chain is one whose states form a single ergodic set; or equivalently a chain in which it is possible to go from every state to every other state, not necessarily in one step.

The existence of steady state conditions in higher order ergodic Markov chains can be demonstrated by computing $P^{(s)}$ for various values of s (see Table I).

Due to the complexity involved in the multiplication of n-dimensional (higher order) matrices, an algorithm is developed in this section to compute the steady state conditions. As in the case of first order Markov chains, steady state probabilities for higher order Markov chains do not depend upon the present and past states of the process.

Ganesan (16) developed the "Reduction technique" to compute the steady state conditions. His technique does not provide a general procedure, but rather a special method for each particular case. The steady state algorithm developed here provides a general method for computing steady state conditions for all cases of higher order chains,

TABLE I

VALUES OF $P^{(s)}$ FOR VARIOUS VALUES OF s

$P^{(1)} =$	0.8000	0.2000	$P^{(6)} =$	0.6280	0.3719	$P^{(11)} =$	0.6033	0.3966
	0.6000	0.4000		0.5770	0.4429		0.5976	0.4023
	0.5000	0.5000		0.6089	0.3910		0.6006	0.3993
	0.3000	0.7000		0.5578	0.4421		0.5949	0.4050
$P^{(2)} =$	0.7500	0.2400	$P^{(7)} =$	0.6178	0.3821	$P^{(12)} =$	0.6022	0.3977
	0.4200	0.5800		0.5884	0.4415		0.5983	0.4016
	0.7000	0.3000		0.6025	0.3974		0.6005	0.3994
	0.3600	0.6400		0.5731	0.4268		0.5966	0.4033
$P^{(3)} =$	0.6920	0.3080	$P^{(8)} =$	0.6120	0.3879	$P^{(13)} =$	0.6014	0.3985
	0.5640	0.4360		0.5908	0.4091		0.5989	0.4010
	0.5900	0.4100		0.6031	0.3968		0.6003	0.3996
	0.4620	0.5380		0.5819	0.4180		0.5978	0.4021
$P^{(4)} =$	0.6664	0.3336	$P^{(9)} =$	0.6077	0.3922	$P^{(14)} =$	0.6009	0.3990
	0.5388	0.4612		0.5947	0.4052		0.5993	0.4006
	0.6280	0.3720		0.6014	0.3985		0.6002	0.3997
	0.5004	0.4996		0.5883	0.41164		0.5985	0.4014
$P^{(5)} =$	0.6408	0.3591	$P^{(10)} =$	0.6051	0.3948	$P^{(15)} =$	0.6006	0.3993
	0.5769	0.4230		0.5961	0.4038		0.5995	0.4004
	0.6026	0.3974		0.6012	0.3987		0.6001	0.3998
	0.5386	0.4613		0.5922	0.4077		0.5990	0.4009

regardless of the number of states. The steady state algorithm was checked against the Reduction technique and was found to yield the same results.

A description of the Reduction technique and the steady state algorithm is presented here, then it will be shown that they yield the same results.

Reduction Technique

The general procedure in this technique is to reduce the transition matrix P of the higher order Markov chain to an equivalent first order matrix. Once a first order matrix is determined, the steady state probabilities are readily obtainable.

The concept of reducing an n-order matrix can be demonstrated with a second order matrix. Let P be the transition probability matrix of a second order Markov chain, then P would appear as given below:

$$P = \begin{bmatrix} p_{111} & p_{112} \\ p_{121} & p_{122} \\ p_{211} & p_{212} \\ p_{221} & p_{222} \end{bmatrix}.$$

P contains two 2 X 2 matrices, one for the state 1 immediately preceding the present state and the other for the state 2. If the process is an ergodic one, these two matrices must be the same at the steady state independent of the states immediately preceding the present state. Since this steady state still depends upon the present state, it is called an intermediate steady state. The intermediate steady state

probabilities can be determined by treating the above matrices as first order ones.

$$\begin{aligned} p^{s}_{121} &= p^{s-1}_{121}\, p_{211} + p^{s-1}_{122}\, p_{221} \\ &= p^{s-1}_{121}\, p_{211} + (1 - p^{s-1}_{121})(1 - p_{222}) \\ &= p^{s-1}_{121}(p_{211} + p_{222} - 1) + (1 - p_{222}) \ . \end{aligned} \tag{3.7}$$

By setting

$$X = p_{211} + p_{222} - 1 \tag{3.8}$$

and

$$Y = (1 - p_{222}) \tag{3.9}$$

then

$$\begin{aligned} p^{s}_{121} &= p^{s-1}_{121}(X) + Y \\ &= (p^{s-2}_{121}(X) + Y)X + Y \\ &= p^{s-2}_{121}(X^2) + Y(1 + X) \\ &= p^{s-3}_{121}(X^3) + Y(1 + X + X^2) \\ &\ \ \vdots \\ &= p_{121}(X^{s-1}) + Y(1 + X + X^2 + \ldots + X^{s-2}) \ . \end{aligned} \tag{3.10}$$

$$\lim_{s \to \infty} p^{s}_{121} = \lim_{s \to \infty} \left[p_{121}(X^{s-1}) + Y(1 + X + X^2 + \ldots + X^{s-2}) \right] \tag{3.11}$$

then

$$\begin{aligned} p^{s}_{121} &= Y(1 - X)^{-1} \\ &= \frac{1 - p_{222}}{1 - p_{211} - p_{222}} \ . \end{aligned} \tag{3.12}$$

Similarly, it can be shown that

$$p^{s}_{121} = p^{s}_{221} = \frac{1 - p_{222}}{2 - p_{211} - p_{222}} \tag{3.13}$$

$$p^{s}_{122} = p^{s}_{222} = \frac{1 - p_{211}}{2 - p_{211} - p_{222}} \tag{3.14}$$

$$p^{s}_{111} = p^{2}_{211} = \frac{1 - p_{122}}{2 - p_{111} - p_{122}} \tag{3.15}$$

$$p^{2}_{122} = p^{2}_{212} = \frac{1 - p_{111}}{2 - p_{111} - p_{122}} \quad . \tag{3.16}$$

It is clear from (3.13), (3.14), (3.15), and (3.16) that at the intermediate steady state the effect upon the process by its state immediately preceding the present state is eliminated and the two 2 X 2 matrices are equal. They would appear as given below:

$$P_1 = \begin{bmatrix} p^{s}_{111} & p^{s}_{122} \\ p^{s}_{121} & p^{s}_{122} \\ p^{s}_{211} & p^{s}_{212} \\ p^{s}_{221} & p^{s}_{222} \end{bmatrix} .$$

If the results of (3.13), (3.14), (3.15), and (3.16) are utilized for P_1, it would appear as given below:

$$P_1 = \begin{bmatrix} p^{s}_{11} & p^{s}_{12} \\ p^{s}_{21} & p^{s}_{22} \\ p^{s}_{11} & p^{s}_{12} \\ p^{s}_{21} & p^{s}_{22} \end{bmatrix} .$$

From the above, it is observed that P_1 consists of two identically equal 2 X 2 first order matrices. The steady state probabilities for these first order matrices can be computed using the same procedure for the determination of intermediate steady state probabilities. At the steady state all the probability vectors will be the same. This is due to the elimination of the effect of the process's present state. The steady state probability matrix would appear as given below:

$$\begin{bmatrix} p_1^s & p_2^s \\ p_1^s & p_2^s \\ p_1^s & p_2^s \\ p_1^s & p_2^s \end{bmatrix}$$

where

$$p_1^s = \frac{1 - p_{22}^s}{2 - p_{11}^s - p_{22}^s} \tag{3.17}$$

$$p_2^s = \frac{1 - p_{11}^s}{2 - p_{11}^s - p_{22}^s} \; . \tag{3.18}$$

Now consider the transition probability matrix for the second order Markov chain given by

$$P = \begin{bmatrix} a & b \\ c & d \\ e & f \\ g & h \end{bmatrix} . \tag{3.19}$$

To compute the steady state conditions substitute in Equations (3.13), (3.14), (3.15), and (3.16) one obtains

$$P_1 = \begin{bmatrix} \frac{1-d}{2-a-d} & \frac{1-a}{2-a-d} \\ \frac{1-h}{2-e-h} & \frac{1-e}{2-e-h} \\ \frac{1-d}{2-a-d} & \frac{1-a}{2-a-d} \\ \frac{1-h}{2-e-h} & \frac{1-e}{2-e-h} \end{bmatrix} = \begin{bmatrix} \frac{d-1}{d-1-b} & \frac{-b}{d-1-b} \\ \frac{-g}{e-1-g} & \frac{e-1}{e-1-g} \\ \frac{d-1}{d-1-b} & \frac{-b}{d-1-b} \\ \frac{-g}{e-1-g} & \frac{e-1}{e-1-g} \end{bmatrix}$$

Substituting back in (3.17) and (3.18) one obtains the steady state conditions:

$$p_1^s = \frac{gb + gc}{2gb + gc + fb} \tag{3.20}$$

$$p_2^s = \frac{gb + fb}{2gb + gc + fb} \quad . \tag{3.21}$$

Steady State Algorithm

Before describing this algorithm, define the following terms:

P_r = The transition probability matrix for the rth order chain where r = 1, 2, ..., k; i.e., P_1 is the transition probability matrix for the 1st order chain, P_2 is the transition probability matrix for the 2nd order chain, and so on.

H_r = Is an r order matrix with some of its elements equal 1 and the rest equal zero. The matrix H_r consists of m^{r-1} two-dimensional matrices of equal size as shown below:

$$H_r = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \\ h_1 \\ h_2 \\ \vdots \\ h_m \\ \vdots \\ h_1 \\ h_2 \\ \vdots \\ h_m \end{bmatrix}, \quad \text{where any } \underset{m \times m}{h_j} = \begin{bmatrix} 00 \ \ldots \ 1 \ \ldots \ 0 \\ 00 \ \ldots \ 1 \ \ldots \ 0 \\ \vdots \quad\quad \vdots \quad\quad \vdots \\ 00 \ \ldots \ 1 \ \ldots \ 0 \end{bmatrix} \leftarrow \text{jth column}$$

i.e., for $r = 3$, $m = 2$, $\therefore m^{r-1} = 4$

$$H_3 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ h_1 \\ h_2 \end{bmatrix}.$$

Next we define the "box" notation $\square$ for matrix multiplication.

$$\text{If} \quad C = \left[c_{i_1 i_2 i_3}\right], \quad A = \left[a_{i_1 i_2 i_3}\right], \quad B = \left[b_{i_1 i_2 i_3}\right] \qquad (3.22)$$

then if $C = A \square B$, one can obtain any element in C as follows:

$$c_{i_1 i_2 i_3} = a_{i_1 i_2 i_3} \cdot b_{i_1 i_2 i_3} \text{ for all } i_1, i_2, i_3 = 1, \ldots, m. \quad (3.23)$$

Algorithm

1 - Set $r = k$.

2 - Compute $B_r = P_r - I$.

3 - Compute $B_r^* = H_r - B_r \square H_r$.

4 - Compute $\left[B_r + B_r^*\right]^{-1}$ which is equal to

$$\begin{bmatrix} P_{r-1} \\ \vdots \\ P_{r-1} \end{bmatrix} .$$

5 - Set $r = r - 1$, proceed to step 6.

6 - If $r = 1$ go to step 7, otherwise go to step 2.

7 - Find a solution for this system of equations:

$$\underset{\sim}{w}' J_1^m = 1$$

$$\underset{\sim}{w}' P_1 = \underset{\sim}{w}'$$

where $\underset{\sim}{w}$ is an $m \times 1$ vector that contains the probabilities which exist at the steady state conditions and J_1^m is a column vector of ones.

Now consider the same matrix P described in (3.19). Using the steady state Algorithm, it will now be shown that the same results can be obtained as with the Reduction technique.

Steps

1. $r = 2$.

2. $B_2 = P_2 - I$

$$B_2 = \begin{bmatrix} a-1 & b \\ c & d-1 \\ e-1 & f \\ g & h-1 \end{bmatrix}.$$

3. $B_2^* = H_2 - B_2 \square H_2$

$$B_2^* = \begin{bmatrix} 2-a & 0 \\ 1-c & 0 \\ 0 & 1-f \\ 0 & 2-h \end{bmatrix}.$$

4.
$$\left[B_2 + B_2^*\right]^{-1} = \begin{bmatrix} \frac{d-1}{d-1-b} & \frac{-b}{d-1-b} \\ \frac{-g}{e-1-g} & \frac{e-1}{e-1-g} \\ \frac{d-1}{d-1-b} & \frac{-b}{d-1-b} \\ \frac{-g}{e-1-g} & \frac{e-1}{e-1-g} \end{bmatrix}.$$

7. Solving the system of equations

$$\underset{\sim}{w}' J_1^2 = 1$$

$$\underset{\sim}{w}' P_1 = \underset{\sim}{w}'$$

one obtains
$$w_1 = \frac{gb + gc}{2gb + gc + fb} \tag{3.24}$$

$$w_2 = \frac{gb + fb}{2gb + gc + fb} \tag{3.25}$$

which is the same results obtained in Equations (3.20) and (3.21).

Example 3-1: Given the following transition probability matrix P for a 3rd order Markov chain having three states, we want to find the steady state probability matrix.

$$P = \begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0.5 & 0.2 & 0.3 \\ 0.4 & 0.1 & 0.5 \\ 0.7 & 0.1 & 0.2 \\ 0.6 & 0.2 & 0.2 \\ 0.2 & 0.6 & 0.2 \\ 0.5 & 0.3 & 0.2 \\ 0.3 & 0.3 & 0.4 \\ 0.2 & 0.3 & 0.5 \\ 0.5 & 0.1 & 0.4 \\ 0.4 & 0.3 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.1 & 0.8 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.3 & 0.2 & 0.5 \\ 0.4 & 0.4 & 0.2 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.3 & 0.4 \\ 0.7 & 0.2 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.2 & 0.7 & 0.1 \\ 0.5 & 0.4 & 0.1 \\ 0.4 & 0.5 & 0.1 \\ 0.6 & 0.1 & 0.3 \\ 0.8 & 0.1 & 0.2 \\ 0.7 & 0.1 & 0.2 \\ 0.2 & 0.7 & 0.1 \end{bmatrix}$$

We follow the steps as they occur in the algorithm:

Step 1: Set $r = k = 3$.

Step 2: Compute $B_3 = P_3 - I$

$$B_3 = P_3 - I = \begin{bmatrix} -0.4 & 0.3 & 0.1 \\ 0.5 & -0.8 & 0.3 \\ 0.4 & 0.1 & -0.5 \\ -0.3 & 0.1 & 0.2 \\ 0.6 & -0.8 & 0.2 \\ 0.2 & 0.6 & -0.8 \\ -0.5 & 0.3 & 0.2 \\ 0.3 & -0.7 & 0.4 \\ 0.2 & 0.3 & -0.5 \\ -0.5 & 0.1 & 0.4 \\ 0.4 & -0.7 & 0.3 \\ 0.2 & 0.5 & -0.7 \\ -0.9 & 0.8 & 0.1 \\ 0.2 & -0.4 & 0.2 \\ 0.3 & 0.2 & -0.5 \\ -0.6 & 0.4 & 0.2 \\ 0.6 & -0.8 & 0.2 \\ 0.3 & 0.3 & -0.6 \\ -0.3 & 0.2 & 0.1 \\ 0.1 & -0.3 & 0.2 \\ 0.2 & 0.7 & -0.9 \\ -0.5 & 0.4 & 0.1 \\ 0.4 & -0.5 & 0.1 \\ 0.6 & 0.1 & -0.7 \\ -0.2 & 0.1 & 0.1 \\ 0.7 & -0.9 & 0.2 \\ 0.2 & 0.7 & -0.9 \end{bmatrix}$$

Step 3: Compute $B_3^* = H_3 - B_3 \square H_3$

$$B_3^* = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} -0.4 & 0.3 & 0.1 \\ 0.5 & -0.8 & 0.3 \\ 0.4 & 0.1 & -0.5 \\ -0.3 & 0.1 & 0.2 \\ 0.6 & -0.8 & 0.2 \\ 0.2 & 0.6 & -0.8 \\ -0.5 & 0.3 & 0.2 \\ 0.3 & -0.7 & 0.4 \\ 0.2 & 0.3 & -0.5 \\ -0.5 & 0.1 & 0.4 \\ 0.4 & -0.7 & 0.3 \\ 0.2 & 0.5 & -0.7 \\ -0.9 & 0.8 & 0.1 \\ 0.2 & -0.4 & 0.2 \\ 0.3 & 0.2 & -0.5 \\ -0.6 & 0.4 & 0.2 \\ 0.6 & -0.8 & 0.2 \\ 0.3 & 0.3 & -0.6 \\ -0.3 & 0.2 & 0.1 \\ 0.1 & -0.3 & 0.2 \\ 0.2 & 0.7 & -0.9 \\ -0.5 & 0.4 & 0.1 \\ 0.4 & -0.5 & 0.1 \\ 0.6 & 0.1 & -0.7 \\ -0.2 & 0.1 & 0.1 \\ 0.7 & -0.9 & 0.2 \\ 0.2 & 0.7 & -0.9 \end{bmatrix} \square \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1.4 & 0 & 0 \\ 0.5 & 0 & 0 \\ 0.6 & 0 & 0 \\ 0 & 0.9 & 0 \\ 0 & 1.8 & 0 \\ 0 & 0.4 & 0 \\ 0 & 0 & 0.8 \\ 0 & 0 & 0.6 \\ 0 & 0 & 1.5 \\ 0.8 & 0 & 0 \\ 1.5 & 0 & 0 \\ 0.6 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & 1.4 & 0 \\ 0 & 0.8 & 0 \\ 0 & 0 & 0.8 \\ 0 & 0 & 0.8 \\ 0 & 0 & 1.6 \\ 1.3 & 0 & 0 \\ 0.9 & 0 & 0 \\ 0.8 & 0 & 0 \\ 0 & 0.6 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0.9 & 0 \\ 0 & 0 & 0.9 \\ 0 & 0 & 0.8 \\ 0 & 0 & 1.9 \end{bmatrix}.$$

Step 4: Compute $\left[B_3 + B_3^*\right]^{-1}$

$$\left[B_3 + B_3^*\right] = \begin{bmatrix}
1 & 0.3 & 0.1 \\
1 & -0.8 & 0.3 \\
1 & 0.1 & -0.5 \\
-0.3 & 1 & 0.2 \\
0.6 & 1 & 0.2 \\
0.2 & 1 & -0.8 \\
-0.5 & 0.3 & 1 \\
0.3 & -0.7 & 1 \\
0.2 & 0.3 & 1 \\
1 & 0.1 & 0.4 \\
1 & -0.7 & 0.3 \\
1 & 0.5 & -0.7 \\
-0.9 & 1 & 0.1 \\
0.2 & 1 & 0.2 \\
0.3 & 1 & -0.5 \\
-0.6 & 0.4 & 1 \\
0.6 & -0.8 & 1 \\
0.3 & 0.3 & 1 \\
1 & 0.2 & 0.1 \\
1 & -0.3 & 0.2 \\
1 & 0.7 & -0.9 \\
-0.5 & 1 & 0.1 \\
0.4 & 1 & 0.1 \\
0.6 & 1 & -0.7 \\
-0.2 & 0.1 & 1 \\
0.7 & -0.9 & 1 \\
0.2 & 0.7 & 1
\end{bmatrix}$$

$$\left[B_3 + B_3^*\right]^{-1} = \begin{bmatrix}
0.530 & 0.229 & 0.241 \\
0.577 & 0.222 & 0.201 \\
0.329 & 0.300 & 0.371 \\
0.370 & 0.293 & 0.337 \\
0.205 & 0.538 & 0.257 \\
0.437 & 0.312 & 0.251 \\
0.288 & 0.556 & 0.156 \\
0.472 & 0.403 & 0.125 \\
0.712 & 0.170 & 0.118 \\
0.530 & 0.229 & 0.241 \\
0.577 & 0.222 & 0.201 \\
0.329 & 0.300 & 0.371 \\
0.370 & 0.293 & 0.337 \\
0.205 & 0.538 & 0.257 \\
0.437 & 0.312 & 0.251 \\
0.288 & 0.556 & 0.156 \\
0.472 & 0.403 & 0.125 \\
0.712 & 0.170 & 0.118 \\
0.530 & 0.229 & 0.241 \\
0.577 & 0.222 & 0.201 \\
0.329 & 0.300 & 0.371 \\
0.370 & 0.293 & 0.337 \\
0.205 & 0.538 & 0.257 \\
0.437 & 0.312 & 0.251 \\
0.288 & 0.556 & 0.156 \\
0.472 & 0.403 & 0.125 \\
0.712 & 0.170 & 0.118
\end{bmatrix}$$

$$\therefore P_2 = \begin{bmatrix} 0.530 & 0.229 & 0.241 \\ 0.577 & 0.222 & 0.201 \\ 0.329 & 0.300 & 0.371 \\ 0.370 & 0.293 & 0.337 \\ 0.205 & 0.538 & 0.257 \\ 0.437 & 0.312 & 0.251 \\ 0.288 & 0.556 & 0.156 \\ 0.472 & 0.403 & 0.125 \\ 0.712 & 0.170 & 0.118 \end{bmatrix} .$$

Step 5: Set r = r - 1 = 3 - 1 = 2, proceed to step 6.

Step 6: $\because$ r = 2 and not 1, go to step 2.

Step 2′: Compute $B_2 = P_2 - I$

$$B_2 = P_2 - I = \begin{bmatrix} -0.470 & 0.229 & 0.241 \\ 0.577 & -0.778 & 0.201 \\ 0.329 & 0.300 & -0.629 \\ -0.630 & 0.293 & 0.337 \\ 0.205 & -0.462 & 0.257 \\ 0.437 & 0.312 & -0.749 \\ -0.712 & 0.556 & 0.156 \\ 0.472 & -0.597 & 0.125 \\ 0.712 & 0.170 & -0.882 \end{bmatrix} .$$

Step 3′: Compute $B_2^* = H_2 - B_2 \square H_2$

$$B_2^* = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} -0.470 & 0.229 & 0.241 \\ 0.577 & -0.778 & 0.201 \\ 0.329 & 0.300 & -0.629 \\ -0.630 & 0.293 & 0.337 \\ 0.205 & -0.462 & 0.257 \\ 0.437 & 0.312 & -0.749 \\ -0.712 & 0.556 & 0.156 \\ 0.472 & -0.597 & 0.125 \\ 0.712 & 0.170 & -0.882 \end{bmatrix} \square \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1.47 & 0 & 0 \\ 0.423 & 0 & 0 \\ 0.671 & 0 & 0 \\ 0 & 0.707 & 0 \\ 0 & 1.462 & 0 \\ 0 & 0.688 & 0 \\ 0 & 0 & 0.844 \\ 0 & 0 & 0.875 \\ 0 & 0 & 1.882 \end{bmatrix} .$$

Step 4′: Compute $\left[B_2 + B_2^*\right]^{-1}$

$$\left[B_2 + B_2^*\right] = \begin{bmatrix} 1 & 0.229 & 0.241 \\ 1 & -0.778 & 0.201 \\ 1 & 0.300 & -0.629 \\ -0.630 & 1 & 0.337 \\ 0.205 & 1 & 0.257 \\ 0.437 & 1 & -0.749 \\ -0.712 & 0.556 & 1 \\ 0.472 & -0.597 & 1 \\ 0.712 & 0.170 & 1 \end{bmatrix} \quad \left[B_2 + B_2^*\right]^{-1} = \begin{bmatrix} 0.488 & 0.246 & 0.266 \\ 0.324 & 0.395 & 0.281 \\ 0.427 & 0.436 & 0.137 \\ 0.488 & 0.246 & 0.266 \\ 0.324 & 0.395 & 0.281 \\ 0.427 & 0.436 & 0.137 \\ 0.488 & 0.246 & 0.266 \\ 0.324 & 0.395 & 0.281 \\ 0.427 & 0.436 & 0.137 \end{bmatrix}.$$

$$\therefore \quad P_1 = \begin{bmatrix} 0.488 & 0.246 & 0.266 \\ 0.324 & 0.395 & 0.281 \\ 0.427 & 0.436 & 0.137 \end{bmatrix}.$$

Step 5′: Set r = r - 1 = 2 - 1 = 1, proceed to step 6.

Step 6′: ∵ r = 1 go to step 7.

Step 7: We find the steady state condition for the 1st order chain whose transition probability matrix is given by P_1: i.e., we solve the problem

$$\underset{\sim}{w}' J_1^1 = 1$$

$$\underset{\sim}{w}' P_1 = \underset{\sim}{w} .$$

We get

$$\underset{\sim}{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 0.315 \\ 0.361 \\ 0.324 \end{bmatrix}. \qquad (3.26)$$

∴ The steady state conditions for the 3rd order Markov chain example is given by (3.26). This example was solved by Ganesan (16) and the same results were obtained.

CHAPTER IV

ANALYSIS OF HIGHER ORDER ABSORBING MARKOV CHAINS

In order to study various aspects of Markov chains it is necessary to recognize several types of state. Then by considering which kinds of state occur in a given chain, the chain may be classified and its special properties discussed. Few definitions are given below, a more detailed description is given by Kemeny and Snell (26).

Definition 4-1: A state i is said to be transient if and only if starting from state i there is a positive probability that the process may not eventually return to this state.

Definition 4-2: A state i is said to be recurrent if and only if, starting from state i, eventual return to this state is certain.

Definition 4-3: An absorbing state is a state which once entered is never left.

Definition 4-4: Absorbing Markov chains are chains all of whose non-transient states are absorbing.

The following quantities are of major interest in the study and analysis of higher order absorbing Markov chains.

1) The mean and variance of the total number of visits from a

transient state i to another transient state j before the process enters an absorbing state.

2) The mean of the total number of steps in transient states before entering a recurrent state.

3) The probability that the process is absorbed in the absorbing state j, given that it started in the transient state i.

Before developing methods for computing these quantities, consider the three-dimensional transitional probability matrix P for a second order absorbing Markov chain. Let P be displayed as a two-dimensional array. The rows of P may be arranged to conform to the configuration

$$P = \begin{array}{c|c} I_1 & O \\ \hline R & Q \end{array}$$

with the indicated partitions described as follows. The partitions I_1 and O are respectively, an array of zero's and ones and an array of all zero's, which have no bearing on the consequent computations. Of importance here are the partitions Q and R. Let r denote the number of absorbing states and (m-r) the number of non-absorbing (transient) states. Then Q is an $(m-r) \times (m-r) \times (m-r)$ matrix with probabilities of transition only among the transient states for its elements. R is an $(m-r) \times (m-r) \times r$ matrix whose elements are the probabilities of the one-step transition from the (m-r) transient states to the r recurrent states.

The Mean of the Total Number of Visits from a Transient State i to Another Transient State j Before the Process Enters a Recurrent State

The expected number of times the process will be in a non-absorbing (transient) state j is given by: Expected number of times in j = (1)(probability of being in j at the start) + (1)(probability of being in j after one step) + (1)(probability of being j after two steps) + ...

The sum of this series is given by M(Fundamental matrix) where

$$M = I + Q + Q^2 + \ldots + Q^n \ . \tag{4.1}$$

This is a geometric series, but unfortunately due to the fact that $MI \neq IM$ and that the Associative law of multiplication does not hold for higher order matrices, then

$$M \neq (I - Q)^{-1} \ . \tag{4.2}$$

The following method is developed to find M. Multiply Equation (4.1) by the matrix Q from the left, one obtains

$$QM = Q + Q^2 + Q^3 + \ldots + Q^n + Q^{n+1} \tag{4.3}$$

$$M - QM = I - Q^{n+1} \ . \tag{4.4}$$

We know that $Q^{n+1} \to 0$ when $n \to \infty$, therefore

$$M - QM = I \ . \tag{4.5}$$

Although the preceding discussion was presented for second order Markov chains, Equation (4.5) holds also for higher order chains, it represents a system of $(m-r)^{k+1}$ equations in $(m-r)^{k+1}$ unknowns, the unknowns being the elements of the matrix M, and k is the order of the chain ($k = 2,3,\ldots$).

One can write this system of linear equations in vector notation

$$Q^* \underset{\sim}{m}^* = \underset{\sim}{I}^* , \tag{4.6}$$

where

Q^* is an $(m-r)^{k+1}$ by $(m-r)^{k+1}$ <u>two-dimensional matrix</u>,

$\underset{\sim}{m}^*$ is an $(m-r)^{k+1} \times 1$ column vector, and

$\underset{\sim}{I}^*$ is an $(m-r)^{k+1} \times 1$ column vector with all the elements either zero or one.

In this system of linear equations (4.6) one will solve for the unknown column vector $\underset{\sim}{m}^* = Q^{*-1} \underset{\sim}{I}^*$. Permuting the rows and columns of the matrix of this system of equations (4.6) one gets the matrix Q^* which can be partitioned into $(m-r)^2$ square submatrices of equal size as follows:

$$\underset{\left[(m-r)^{k+1} \times (m-r)^{k+1}\right]}{Q^*} = \begin{bmatrix} Q^*_{11} & Q^*_{12} & \cdots & Q^*_{1(m-r)} \\ Q^*_{21} & Q^*_{22} & \cdots & Q^*_{2(m-r)} \\ \vdots & \vdots & & \vdots \\ Q^*_{(m-r)1} & Q^*_{(m-r)2} & \cdots & Q^*_{(m-r)(m-r)} \end{bmatrix}.$$

Noting that

$$Q^*_{ij} = 0 \qquad \text{for} \qquad i \neq j$$

and

$$Q^*_{11} = Q^*_{22} = \cdots = Q^*_{(m-r)(m-r)} \neq 0$$

and

Q^*_{ij} is an $\left[(m-r)^k \times (m-r)^k\right]$ square matrix for $i,j = 1,\ldots(m-r)$,

then

$$\underset{\left[(m-r)^{k+1} \times (m-r)^{k+1}\right]}{Q^{*-1}} = \begin{bmatrix} Q^{*-1}_{ii} & & & \\ & Q^{*-1}_{ii} & & \\ & & \ddots & \\ & & & Q^{*-1}_{ii} \end{bmatrix}.$$

So the problem now reduces to finding Q_{ii}^{*-1}. Hence

$$\underset{\sim}{m}^{*} = Q^{*-1} \underset{\sim}{I}^{*} . \qquad (4.7)$$

In many actual problems the matrix Q_{ii}^{*} would be of large size but with many zero coefficients. Such a matrix is known as a sparse matrix. The problem of finding the inverse of a sparse matrix have been studied and reviewed in detail by Tewarson (41), Matsushita (33), Walsh (43), Nathan and Even (35), and many others.

Example 4-1

Consider a second order Markov chain with m = 3. The transition matrix P is given below.

$$\underset{(3 \times 3 \times 3)}{P} = \begin{bmatrix} 0.3 & 0.3 & 0.4 \\ 0.2 & 0.2 & 0.6 \\ 0 & 0 & 1 \\ 0.5 & 0.4 & 0.1 \\ 0.4 & 0.2 & 0.4 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} .$$

For the above absorbing state 3, with m = 3, k = 2, r = 1, and (m-r) = 2. The matrices Q and R are given below:

$$\underset{(2 \times 2 \times 2)}{Q} = \begin{bmatrix} 0.3 & 0.3 \\ 0.2 & 0.2 \\ 0.5 & 0.4 \\ 0.4 & 0.2 \end{bmatrix}, \qquad \underset{(2 \times 2 \times 1)}{R} = \begin{bmatrix} 0.4 \\ 0.6 \\ 0.1 \\ 0.4 \end{bmatrix} .$$

$$\therefore \quad M - QM = I$$

$$\begin{bmatrix} m_1 & m_5 \\ m_2 & m_6 \\ m_3 & m_7 \\ m_4 & m_8 \end{bmatrix} - \begin{bmatrix} 0.3 & 0.3 \\ 0.2 & 0.2 \\ 0.5 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} \begin{bmatrix} m_1 & m_5 \\ m_2 & m_6 \\ m_3 & m_7 \\ m_4 & m_8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

This is a system of $(m-r)^{k+1} = (3-1)^{2+1} = 8$ equations in eight unknowns, which can be written in vector notations as

$$Q^* \underset{\sim}{m}^* = \underset{\sim}{I}^*$$

$$\begin{bmatrix} 0.7 & -0.3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.7 & -0.3 & 0 & 0 \\ 0 & 1 & -0.2 & -0.2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -0.2 & -0.2 \\ -0.5 & -0.4 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.5 & -0.4 & 1 & 0 \\ 0 & 0 & -0.4 & 0.8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.4 & 0.8 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \\ m_6 \\ m_7 \\ m_8 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Permuting the rows, one obtains:

$$\begin{bmatrix} 0.7 & -0.3 & & & & & & \\ & 1 & -0.2 & -0.2 & & & & \\ -0.5 & -0.4 & 1 & & & & & \\ & & -0.4 & 0.8 & & & & \\ & & & & 0.7 & -0.3 & & \\ & & & & & 1 & -0.2 & -0.2 \\ & & & & -0.5 & -0.4 & 1 & \\ & & & & & & -0.4 & 0.8 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \\ m_6 \\ m_7 \\ m_8 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

$$\therefore\ Q^{*}_{ii} = \begin{bmatrix} 0.7 & -0.3 & 0 & 0 \\ 0 & 1 & -0.2 & -0.2 \\ -0.5 & -0.4 & 1 & 0 \\ 0 & 0 & -0.4 & 0.8 \end{bmatrix},\quad Q^{*-1}_{ii} = \begin{bmatrix} 1.54 & 0.525 & 0.157 & 0.131 \\ 0.263 & 1.226 & 0.367 & 0.312 \\ 0.876 & 0.753 & 1.226 & 0.188 \\ 0.438 & 0.376 & 0.613 & 1.344 \end{bmatrix}$$

$$\underset{\sim}{m}^{*} = Q^{*-1}\,\underset{\sim}{I}^{*}$$

$$= \begin{bmatrix} 1.54 & 0.525 & 0.157 & 0.131 & & & & \\ 0.263 & 1.226 & 0.367 & 0.312 & & & & \\ 0.876 & 0.753 & 1.226 & 0.188 & & & & \\ 0.438 & 0.376 & 0.613 & 1.344 & & & & \\ & & & & 1.54 & 0.525 & 0.157 & 0.131 \\ & & & & 0.263 & 1.226 & 0.367 & 0.312 \\ & & & & 0.876 & 0.753 & 1.226 & 0.188 \\ & & & & 0.438 & 0.376 & 0.613 & 1.344 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.697 \\ 0.630 \\ 2.102 \\ 1.051 \\ 0.656 \\ 1.538 \\ 0.941 \\ 1.72 \end{bmatrix}.$$

$$\therefore\ M = \begin{bmatrix} 1.697 & 0.656 \\ 0.630 & 1.538 \\ 2.102 & 0.941 \\ 1.051 & 1.72 \end{bmatrix}.$$

Let the m state second order Markov chain consist of r recurrent states and (m-r) transient states. Let T be the set of these transient states and T^{c} the set of recurrent states.

Let $N_{hij}(h,i,j \in T)$ be the random variable denoting the number of times the process visits j before it eventually enters a recurrent state, having initially started from state h and was in i one step later.

Let $M_{hij} = E\left[N_{hij}\right]$.

Theorem 4-1: For $h, i, j \in T$

$$\|M_{hij}\| = M \quad . \tag{4.8}$$

Proof:

Initially the Markov chain was in state h then i one step later, where $h,i \in T$. Now if in one step it enters a recurrent state (with probability $\sum_{r \in T^c} P_{hir}$), the number of visits to j is zero unless $j = i$. If δ_{hij} is a function such that $\delta_{hij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$, then we can write $N_{hij} = \delta_{hij}$ with probability $\sum_{r \in T^c} P_{hir}$.

On the other hand, suppose the Markov chain moves to a state $r \in T$ at the first step (with probability P_{hir}). From that position onward the number of visits to j is N_{irj}. However, if i=j then the total number of visits to j would be $N_{irj} + \delta_{hij}$. Thus we have

$$N_{hij} = \begin{cases} \delta_{hij} & \text{with probability } \sum_{r \in T^c} P_{hir} \\ N_{irj} + \delta_{hij} & \text{with probability } P_{hir},\ r \in T. \end{cases} \tag{4.9}$$

Taking expectations, we get

$$E(N_{hij}) = \sum_{r \in T^c} P_{hir}\, \delta_{hij} + \sum_{r \in T} P_{hir}\, E(N_{irj} + \delta_{hij})$$

$$= \delta_{hij} + \sum_{r \in T} P_{hir}\, M_{irj} \tag{4.10}$$

$$\|M_{hij}\| = I + Q \|M_{hij}\| \tag{4.11}$$

$$M - QM = I \quad .$$

This is the same Equation (4.5) derived earlier. It should be noted the number of visits here is counted with reference to state i. It is apparent that this counting could have begun at state h, in which case

Equation (4.11) becomes

$$M - QM = C \ , \tag{4.12}$$

where

$$C = \| C_{hij} \| \ , \ C_{hij} = \begin{cases} 2 \text{ if } h = i = j \\ 1 \text{ if } h = j \text{ or } i = j \\ 0 \text{ if } h \neq i \neq j \end{cases} .$$

The Variance of the Number of Visits from a Transient State i to Another Transient State j Before the Process Enters a Recurrent State

Related to the matrix M, define the matrices M_d, M_2.

Note: if $M = \begin{bmatrix} a & b \\ c & d \\ e & f \\ g & h \end{bmatrix}$, then $M_d = \begin{bmatrix} a & o \\ o & d \\ e & o \\ o & h \end{bmatrix}$, and $M_2 = \begin{bmatrix} a^2 & b^2 \\ c^2 & d^2 \\ e^2 & f^2 \\ g^2 & h^2 \end{bmatrix}$.

Let $\sigma^2_{hij} = V(N_{hij})$.

Theorem 4-2: For h, i, j $\in$ T

$$\| \sigma^2_{hij} \| = M_1 - M_2 \ , \tag{4.13}$$

where M_1 is the solution to this system of equations

$$\| E(N^2_{hij}) \| - Q \| E(N^2_{hij}) \| = 2M_d - I \ . \tag{4.14}$$

Proof: By definition

$$\sigma^2_{hij} = V(N_{hij}) = E(N^2_{hij}) - \left[E(N_{hij})\right]^2 . \tag{4.15}$$

We know that $\left[E(N_{hij})\right]^2 = M^2_{hij}$,

hence

$$\left\| \left[E(N_{hij}) \right]^2 \right\| = M_2 \ . \tag{4.16}$$

One can also write

$$N_{hij}^2 = \begin{cases} \delta_{hij}^2 & \text{with probability } \sum_{r \in T^c} P_{hir} \\ (N_{irj}+\delta_{hij})^2 & \text{with probability } P_{hir},\ r \in T. \end{cases} \tag{4.17}$$

One also has $\delta_{hij}^2 = \delta_{hij}$. Taking expectations one obtains

$$E(N_{hij}^2) = \sum_{r \in T^c} P_{hir}\, \delta_{hij}^2 + \sum_{r \in T} P_{hir}\, E(N_{irj} + \delta_{hij})^2$$

$$= \sum_{r \in T^c} P_{hir}\, \delta_{hij}^2 + \sum_{r \in T} P_{hir}\, E(N_{irj}^2) + 2 \sum_{r \in T} P_{hir}\, E(N_{irj})\delta_{hij}$$

$$+ \sum_{r \in T} P_{hir}\, \delta_{hij}^2$$

$$= \delta_{hij} + \sum_{r \in T} P_{hir}\, E(N_{irj}^2) + 2 \sum_{r \in T} P_{hir}\, E(N_{irj})\delta_{hij} \tag{4.18}$$

$$\therefore \left\| E(N_{hij}^2) \right\| = Q \left\| E(N_{hij}^2) \right\| + 2(QM)_d + I \ . \tag{4.19}$$

$$\because \quad M - QM = I$$

$$\therefore \quad QM = M - I$$

$$\therefore \quad (QM)_d = (M-I)_d$$

$$(QM)_d = M_d - I \ . \tag{4.20}$$

Substituting back in (4.19) one obtains

$$\left\| E(N_{hij}^2) \right\| - Q \left\| E(N_{hij}^2) \right\| = 2M_d - I \ . \tag{4.21}$$

Solving this system we get M_1 and hence

$$\| \sigma^2_{hij} \| = M_1 - M_2 \quad .$$

Example 4-2

From the previous example

$$M = \begin{bmatrix} 1.697 & 0.656 \\ 0.630 & 1.538 \\ 2.102 & 0.941 \\ 1.051 & 1.72 \end{bmatrix}$$

$$\therefore M_d = \begin{bmatrix} 1.697 & 0 \\ 0 & 1.538 \\ 2.102 & 0 \\ 0 & 1.72 \end{bmatrix} \quad \text{and} \quad 2M_d = \begin{bmatrix} 3.394 & 0 \\ 0 & 3.076 \\ 4.204 & 0 \\ 0 & 3.44 \end{bmatrix} .$$

$$2M_d - I = \begin{bmatrix} 2.394 & 0 \\ 0 & 2.076 \\ 3.204 & 0 \\ 0 & 2.44 \end{bmatrix} .$$

Now writing Equation (4.14) in vector notation one obtains

$$\underset{\sim}{m}^*_1 = Q^{*-1}(2\underset{\sim}{m}^*_d - \underset{\sim}{I}^*)$$

$$= \begin{bmatrix} 1.54 & 0.525 & 0.157 & 0.131 & & & & \\ 0.263 & 1.226 & 0.367 & 0.312 & & & & \\ 0.876 & 0.753 & 1.226 & 0.188 & & & & \\ 0.438 & 0.376 & 0.613 & 1.344 & & & & \\ & & & & 1.54 & 0.525 & 0.157 & 0.131 \\ & & & & 0.263 & 1.226 & 0.367 & 0.312 \\ & & & & 0.876 & 0.753 & 1.226 & 0.188 \\ & & & & 0.438 & 0.376 & 0.613 & 1.344 \end{bmatrix} \begin{bmatrix} 2.394 \\ 0 \\ 3.204 \\ 0 \\ 0 \\ 2.076 \\ 0 \\ 2.44 \end{bmatrix} = \begin{bmatrix} 4.19 \\ 1.805 \\ 6.025 \\ 3.012 \\ 1.408 \\ 3.306 \\ 2.021 \\ 4.06 \end{bmatrix}$$

$$\therefore \quad M_1 = \begin{bmatrix} 4.19 & 1.408 \\ 1.805 & 3.306 \\ 6.025 & 2.021 \\ 3.012 & 4.06 \end{bmatrix}.$$

By squaring the elements of M one gets

$$M_2 = \begin{bmatrix} 2.879 & 0.430 \\ 0.397 & 2.365 \\ 4.418 & 0.885 \\ 1.05 & 2.958 \end{bmatrix}$$

$$\therefore \quad \| \sigma^2_{hij} \| = M_1 - M_2 = \begin{bmatrix} 1.311 & 0.978 \\ 1.408 & 0.941 \\ 1.607 & 1.36 \\ 1.962 & 1.102 \end{bmatrix}.$$

The Mean of the Total Number of Steps in Transient States Before First Entering a Recurrent State

The variable to be considered is

$$N_{hi} = \sum_{j \in T} N_{hij}.$$

Using the results given by Theorem 4-1, we can get the mean of $N_{hi}(h, i \in T)$. Define

$$M_P = \| \sum_{j \in T} M_{hij} \| \tag{4.22}$$

a column vector where the kth component is the sum of the elements in the kth row of M.

Theorem 4-3

$$\| E(N_{hi}) \| = M_P \ . \tag{4.23}$$

Proof

$$E(N_{hi}) = E\Big(\sum_{j \in T} N_{hij} \Big)$$

$$= \sum_{j \in T} E(N_{hij}) \ .$$

Example 4-3: From Example 4-1

$$M = \begin{bmatrix} 1.697 & 0.656 \\ 0.630 & 1.538 \\ 2.102 & 0.941 \\ 1.051 & 1.72 \end{bmatrix}$$

$$\therefore \quad M_P = \begin{bmatrix} 2.353 \\ 2.168 \\ 3.043 \\ 2.771 \end{bmatrix} \ .$$

The Probability of Absorption by Any Given Absorbing State

In a similar manner the probability of absorption by any given absorbing state is found. Let j signify some given absorbing state; let i signify some specified non-absorbing state.

The probability of ending in j = (probability of going from i to j in 1 step) + (probability of going from i to j in 2 steps) + ...

In the first order Markov chains this can be obtained from the series

$$R + QR + Q^2R + \ldots = (I-Q)^{-1} R \ .$$

Unfortunately this is not the case in higher order Markov chains. According to the properties of multiplication in n-dimensional matrix algebra the probability of ending in $j = R + QR + Q(QR) + Q(Q(QR)) + \ldots$. Let the sum of this series be denoted by B.

$$\therefore \quad B = R + QR + Q(QR) + Q(Q(QR)) + \ldots Q(\ldots(QR)))) \ . \qquad (4.24)$$

Multiplying Equation (4.24) by the matrix Q from the left hand side, one obtains

$$QB = QR + Q(QR) + Q(Q(QR)) + \ldots + Q(Q(\ldots(QR))\ldots) \qquad (4.25)$$

$$B - QB = R - Q(Q(\ldots(QR))\ldots) \ ,$$

but $Q(Q(\ldots(QR))\ldots) \rightarrow 0$ when the order of multiplication gets bigger,

$$\therefore \qquad B - QB = R \ . \qquad (4.26)$$

Writing this in vector notation, one obtains

$$Q^*_{ii} \underset{\sim}{b}^* = \underset{\sim}{R}^*$$

where Q^*_{ii} is the same two-dimensional matrix defined earlier

$\underset{\sim}{b}^*$ is a $(m-r)^k \times 1$ column vector

$\underset{\sim}{R}^*$ is a $(m-r)^k \times 1$ column vector .

Solve for $\underset{\sim}{b}^* = Q^{*-1}_{ii} \underset{\sim}{R}^*$ where Q^{*-1}_{ii} has already been computed.

Example 4-4:

$$\underset{\sim}{b}^{*} = Q_{ii}^{*-1}\ \underset{\sim}{R}^{*}$$

$$= \begin{bmatrix} 1.54 & 0.525 & 0.157 & 0.131 \\ 0.263 & 1.226 & 0.367 & 0.312 \\ 0.876 & 0.753 & 1.226 & 0.188 \\ 0.438 & 0.376 & 0.613 & 1.344 \end{bmatrix} \begin{bmatrix} 0.4 \\ 0.6 \\ 0.1 \\ 0.4 \end{bmatrix} = \begin{bmatrix} 0.9991 \\ 1.0023 \\ 1.000 \\ 0.9997 \end{bmatrix}$$

$$\therefore \quad B = \begin{bmatrix} 1.000 \\ 1.000 \\ 1.000 \\ 1.000 \end{bmatrix} .$$

CHAPTER V

ESTIMATION OF TRANSITION PROBABILITIES OF HIGHER ORDER MARKOV CHAINS

The problem of estimation of transition probabilities for first order Markov chains has received a considerable amount of attention in the literature. This problem has been studied and reviewed in detail by Lee, Judge and Zellner (31), Miller (34), Whittle (44), Lee, Judge and Takayama (32), and others. Both Telser (40) and Anderson and Goodman (1), have studied the problem of estimation of transition probabilities for first order as well as higher order Markov chains.

In this chapter the Maximum Likelihood method of estimation of transition probabilities of higher order Markov chains is presented. In addition, a discussion of the problem of testing the hypothesis that the chain is of a given order is considered.

Maximum Likelihood Estimates of Transition Probabilities

The multinomial distribution is given below since it will be used in the following discussion.

$$p(k_1,k_2,\ldots,k_n) = P(X_1 = k_1, \ldots, X_n = k_n)$$

$$= \frac{k!}{k_1!k_2!\ldots,k_n!k_{n+1}!}\varphi_1^{k_1}\varphi_2^{k_2}\ldots\varphi_{n+1}^{k_{n+1}}$$

where

$$k_{n+1} = k - k_1 - k_2 - \ldots - k_n$$

$$\varphi_{n+1} = 1 - \varphi_1 - \varphi_2 - \ldots - \varphi_n \text{ with } 0 \leq \varphi_r \leq 1$$

$$\text{for } r = 1, \ldots, n+1 \quad . \qquad (5.1)$$

The method of estimation presented in this section will be for second order Markov chains. For higher orders the same procedure is used.

Suppose a second order Markov chain has been observed for n time points (at a stretch) and let n_{sij} be the number of transitions; i.e., the number of individuals in state s at time t-2, in i at time t-1 and in j at time t. These transition counts may be represented as

t-2	t-1	t				
		1	2	...	m	
1	1	n_{111}	n_{112}	...	n_{11m}	n_{11}
	2	n_{121}	n_{122}	...	n_{12m}	n_{12}
	⋮	⋮	⋮		⋮	⋮
	m	n_{1m1}	n_{1m2}	...	n_{1mm}	n_{1m}
2	1	n_{211}	n_{212}	...	n_{21m}	n_{21}
	2	n_{221}	n_{222}	...	n_{22m}	n_{22}
	⋮	⋮	⋮		⋮	⋮
⋮	m	n_{2m1}	n_{2m2}	...	n_{2mm}	n_{2m}
	⋮	⋮	⋮		⋮	⋮
m	1	n_{m11}	n_{m12}	...	n_{m1m}	n_{m1}
	2	n_{m21}	n_{m22}	...	n_{m2m}	n_{m2}
	⋮	⋮	⋮		⋮	⋮
	m	n_{mm1}	n_{mm2}	...	n_{mmm}	n_{mm}
						n

(5.2)

where m is the number of states and

$$n = \sum_{s=1}^{m} \sum_{i=1}^{m} \sum_{j=1}^{m} n_{sij} \text{ and } n_{si} = \sum_{j=1}^{m} n_{sij} .$$

Let the transition probability matrix of the second order Markov chain be $P = [p_{sij}]$. We are interested in the estimates of the elements p_{sij}; we shall denote these estimates by $\hat{p}_{sij}$ $(s,i,j=1,\ldots,m)$.

For a given initial state and a number of trials n_{si}, the sample of transition counts $(n_{si1},n_{si2},\ldots,n_{sim})$ can be considered as a sample of size n_{si} from a multinomial distribution with probabilities $(p_{si1}, p_{si2},\ldots,p_{sim})$ such that $\sum_{j=1}^{m} P_{sij} = 1$. The probability of this outcome can therefore be given as

$$\frac{n_{si}!}{n_{si1}!n_{si2}!\ldots n_{sim}!} p_{si1}^{n_{si1}} p_{si2}^{n_{si2}} \cdots p_{sim}^{n_{sim}} . \qquad (5.3)$$

Extending this argument, the probability of the realization of transition counts as in (5.2) is given by

$$\prod_{s=1}^{m} \prod_{i=1}^{m} \frac{n_{si}!}{n_{si1}!n_{si2}!\ldots n_{sim}!} p_{si1}^{n_{si1}} p_{si2}^{n_{si2}} \cdots p_{sim}^{n_{sim}} . \qquad (5.4)$$

In (5.2) the row sums $(n_{11}, n_{12}, \ldots, n_{mm})$ are also random variables, and therefore the unconditional likelihood function $f(p_{sij})$ of the sample observation consists of another factor giving the joint distribution of these random variables. Whittle (44) has shown that this distribution is independent of the probability elements p_{sij}. Denoting $A(n_{sij})$ to be the contribution of the distribution of $n_{11}, n_{12}, \ldots, n_{mm}$ to the likelihood function, then $f(p_{sij})$ is given by

$$f(p_{sij}) = A(n_{sij}) \prod_{s=1}^{m} \prod_{i=1}^{m} \frac{n_{si}!}{n_{si1}!n_{si2}!\ldots n_{sim}!} p_{si1}^{n_{si1}} p_{si2}^{n_{si2}} \cdots p_{sim}^{n_{sim}} . \qquad (5.5)$$

Taking the natural logarithm, we can write

$$L(p_{sij}) = \ell n B(n_{sij}) + \sum_{s=1}^{m} \sum_{i=1}^{m} \sum_{j=1}^{m} n_{sij} \ell n P_{sij} , \tag{5.6}$$

where $\ell n B(n_{sij})$ contains all terms independent of the p_{sij}'s.

To derive maximum likelihood estimates, we maximize (5.6) under the condition $\sum_{j=1}^{m} p_{sij} = 1$ for $(s,i=1,\ldots,m)$. Incorporating this condition into (5.6), one can write

$$L(p_{sij}) = \ell n B(n_{sij}) + \sum_{s=1}^{m} \sum_{i=1}^{m} \sum_{j=1}^{m} n_{sij} \ell n p_{sij} + \sum_{s=1}^{m} \sum_{i=1}^{m} n_{sim} \ell n (1 - p_{si1} - p_{si2} - \cdots p_{si(m-1)}) . \tag{5.7}$$

From the structure of the log likelihood function $L(P_{sij})$, it is clear that the estimates can be obtained separately for the m values of $s,i=1,2,\ldots,m$. For a specific value of s and i, we have

$$L_{si}(P_{sij}) = \ell n B(n_{sij}) + \sum_{j=1}^{m-1} n_{sij} \ell n p_{sij} + n_{sim} \ell n (1 - p_{si1} - p_{si2} - \cdots - p_{si(m-1)}) . \tag{5.8}$$

Differentiating (5.8) with respect to $P_{sij}(j=1,\ldots,m-1)$ and setting it equal to zero, we get

$$\begin{aligned}
&\frac{n_{si1}}{p_{si1}} - \frac{n_{sim}}{1 - p_{si1} - p_{si2} - \cdots - p_{si(m-1)}} = 0 \\
&\frac{n_{si2}}{p_{si2}} - \frac{n_{sim}}{1 - p_{si1} - p_{si2} - \cdots - p_{si(m-1)}} = 0 \\
&\quad\vdots \qquad\qquad\qquad\qquad \vdots \\
&\frac{n_{si(m-1)}}{p_{si(m-1)}} - \frac{n_{sim}}{1 - p_{si1} - p_{si2} - \cdots - p_{si(m-1)}} = 0 .
\end{aligned} \tag{5.9}$$

Combining these equations, one may write

$$\frac{n_{si1}}{p_{si1}} = \frac{n_{si2}}{p_{si2}} = \cdots = \frac{n_{si(m-1)}}{p_{si(m-1)}} = \frac{n_{sim}}{1 - p_{si1} - p_{si2} - \cdots - p_{si(m-1)}} \cdot \quad (5.10)$$

This leads to

$$\frac{n_{si1}}{n_{si1}} \cdot P_{si1} = P_{si1}$$

$$\frac{n_{si2}}{n_{si1}} \cdot P_{si1} = P_{si2}$$

$$\vdots$$

$$\frac{n_{sim}}{n_{si1}} \cdot P_{si1} = 1 - P_{si1} - p_{si2} - \cdots - P_{si(m-1)} \cdot \quad (5.11)$$

Adding both sides of the equations in (5.11) one obtains

$$\frac{n_{si1} + n_{si2} + \cdots + n_{sim}}{n_{si1}} \cdot p_{si1} = 1 \quad (5.12)$$

which yields the estimate

$$\hat{p}_{si1} = \frac{n_{si1}}{n_{si}} \cdot \quad (5.13)$$

In a similar manner, one can derive estimates of other elements. Thus one obtains

$$\hat{p}_{sij} = \frac{n_{sij}}{n_{si}} \qquad s,i,j = 1,2,\ldots,m \quad (5.14)$$

which is the same result obtained by Anderson and Goodman (1).

Test of the Hypothesis that the Chain is of a Given Order

Anderson and Goodman (1) considered this problem of testing the hypothesis that the chain is of a given order.

They considered testing the null hypothesis that the chain is first-order against the alternative that it is second-order. The null hypothesis is that

$$p_{1ij} = p_{2ij} = \cdots = p_{mij} = p_{ij}, \quad \text{for } i,j = 1,\ldots,m \quad . \tag{5.15}$$

The likelihood ratio criterion for testing this hypothesis is given by Anderson and Goodman (1) as

$$\lambda = \prod_{s,i,j=1}^{m} \left(\frac{\hat{p}_{ij}}{\hat{p}_{sij}} \right)^{n_{sij}} \tag{5.16}$$

where $\hat{p}_{sij}$ is given by (5.14), and

$$\hat{p}_{ij} = \sum_{s=1}^{m} n_{sij} \Big/ \sum_{s=1}^{m} \sum_{\ell=1}^{m} n_{si\ell} \tag{5.17}$$

is the maximum likelihood estimate of $\hat{p}_{ij}$. Under the null hypothesis, Anderson and Goodman (1) showed that $-2\ell n\lambda$ has an asymptotic χ^2 distribution with $m(m-1)^2$ degrees of freedom. The preceding analysis can be directly generalized for a chain of order k. This means that one can test the hypothesis that the process is a chain of order (k-1) against the alternative hypothesis that it is not a (k-1) but a k order chain.

Example 5-1

Consider the following table which gives the transition counts. The maximum likelihood estimate of the transition probabilities will be calculated. Also, the problem of testing the null hypothesis that the chain is first-order against the alternative that it is second-order is considered.

t-2	t-1		t		
	1	125	5	16	146
1	2	7	106	15	128
	3	11	18	142	171
	1	146	2	4	152
2	2	16	111	4	131
	3	40	36	96	172
	1	80	20	20	120
3	2	30	70	29	129
	3	21	65	65	151
					1300

Using Equation (5.14) to estimate p_{111}, one obtains

$$\hat{p}_{111} = \frac{125}{146} = 0.856 \ .$$

Similarly, one obtains the rest of the elements of the matrix

$$\hat{P} = [\hat{p}_{sij}] \ .$$

$$\hat{P} = \begin{bmatrix} 0.856 & 0.034 & 0.110 \\ 0.055 & 0.828 & 0.117 \\ 0.064 & 0.105 & 0.831 \\ 0.961 & 0.013 & 0.026 \\ 0.122 & 0.847 & 0.031 \\ 0.233 & 0.209 & 0.558 \\ 0.667 & 0.167 & 0.166 \\ 0.023 & 0.543 & 0.224 \\ 0.140 & 0.430 & 0.430 \end{bmatrix}$$

To test the null hypothesis that

$$p_{1ij} = p_{2ij} = p_{3ij} = p_{ij}, \qquad \text{for } i,j = 1,2,3$$

one needs to first calculate $\hat{p}_{ij}$ for $i,j = 1,2,3$. Using Equation (5.17) one obtains

$$\hat{p}_{11} = \frac{125 + 146 + 80}{146 + 152 + 120} = 0.839.$$

Similarly,

$$\hat{p}_{12} = 0.065, \quad \hat{p}_{13} = 0.096, \quad \hat{p}_{21} = 0.136, \quad \hat{p}_{22} = 0.740$$

$$\hat{p}_{23} = 0.124, \quad \hat{p}_{31} = 0.146, \quad \hat{p}_{32} = 0.240, \quad \hat{p}_{33} = 0.614 .$$

Using Equation (5.16), and taking the natural logarithm one obtains

$$\ell n\lambda = \sum_{s=1}^{3}\sum_{i=1}^{3}\sum_{j=1}^{3} n_{sij}\,[\ell n\hat{p}_{ij} - \ell n\hat{p}_{sij}] .$$

$$\therefore \quad -2\ell n\lambda = \sum_{s=1}^{3}\sum_{i=1}^{3}\sum_{j=1}^{3} 2n_{sij}[\ell n\hat{p}_{sij} - \ell n\hat{p}_{ij}]$$

$$= 2[125(0.155-0.175) + 5(3.38-2.73) + \ldots + 65(0.843-1.427) + 65(0.844-0.487)]$$

$$= 283.22 .$$

The degrees of freedom is $m(m-1)^2 = 12$.

From χ^2 tables, we find

$$P(\chi^2 \geq 283.22) < 0.0001 .$$

This shows that we can reject the hypothesis that the chain is first order even with 0.01% significance level.

CHAPTER VI

HIGHER ORDER SEMI-MARKOV CHAINS

First order discrete-time semi-Markov chains have been studied and reviewed in detail by Howard (23). A short discussion of such chains will follow.

A first order discrete-time semi-Markov chain is a process whose successive state occupancies are governed by the transition probabilities of a Markov process, but whose stay in any state is described by an integer-valued random variable that depends on the state presently occupied and on the state to which the next transition will be made.

Let p_{ij} be the probability that a semi-Markov process that entered state i on its last transition will enter state j on its next transition. p_{ij} must satisfy the following:

$$p_{ij} \geq 0 \quad , \quad i,j = 1,\ldots,m \tag{6.1}$$

$$\sum_{j=1}^{m} p_{ij} = 1 \quad , \quad i = 1, 2, \ldots, m \tag{6.2}$$

where m is the total number of states in the system. Whenever a process enters a state i, we imagine that it determines the next state j to which it will move according to state i's transition probabilities P_{i1}, P_{i2}, ..., P_{im}. However, after j has been selected, but before making this transition from state i to state j, the process "holds" for a time in state i. The holding times are positive, integer valued random

variable each governed by a probability mass function called the holding time mass function for a transition from state i to state j.

For a first order discrete-time semi-Markov process to be described completely, we must specify m^2 holding time mass functions, in addition to the transition probabilities. Figures 5 and 6 better describe the difference between a first order Markov process and a first order semi-Markov process.

One can consider a first order discrete Markov process to be a special case of a first order semi-Markov process.

Higher Order Discrete-Time Semi-Markov Chains

The analysis for second order chains is presented in this section. For orders higher than two, the analysis follow in a similar manner.

The second order discrete-time semi-Markov process is a process whose successive state occupancies are governed by the transition probabilities of a second order Markov chain, but whose stay in any state is described by an integer-valued random variable that depends on the state presently occupied, the previous occupied state, and on the state to which the next transition will be made.

Define the following variables:

m = Number of states in the process.

p_{sij} = The probability that a semi-Markov process that <u>entered</u> state i on its last transition will <u>enter</u> state j on its next transition given that it <u>entered</u> state s time before last.

τ_{sij} = Holding time. The time the process will spend in state i

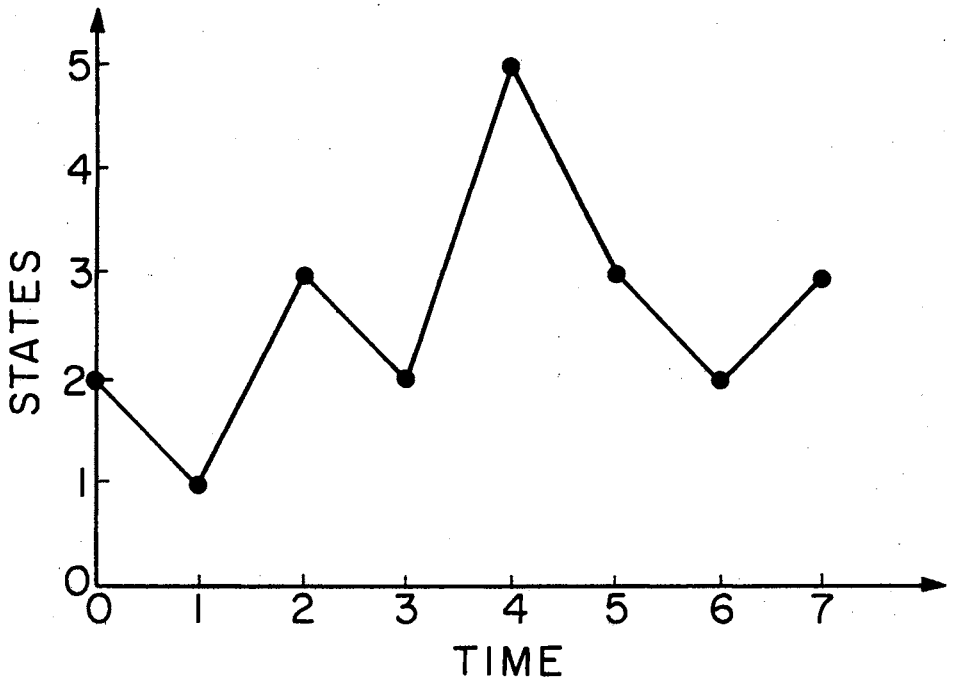

Figure 5. A Possible Markov Process Trajectory

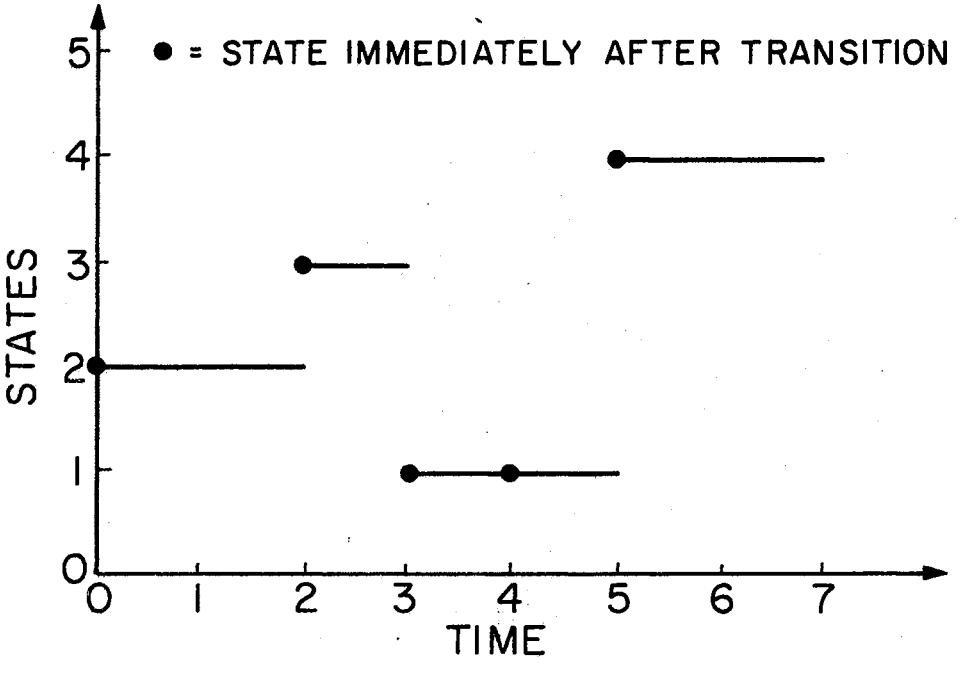

Figure 6. A Possible Semi-Markov Process Trajectory

now before making a transition to state j given that it was in state s last time.

$h_{sij}(\cdot)$ = The probability mass function for τ_{sij}, the holding time.

τ_{si} = The waiting time for state i, the time the process will spend in state i before making a transition given that it was in state s last time.

$w_{si}(\cdot)$ = The probability mass function for τ_{si}, the waiting time.

The transition probabilities must satisfy the following equations.

$$P_{sij} \geq 0 \qquad s = 1,2,\ldots,m \quad i = 1,2,\ldots,m \quad j = 1,2,\ldots,m \tag{6.3}$$

$$\sum_{j}^{m} P_{sij} = 1 \qquad i = 1,\ldots,m \quad s = 1,\ldots,m \; . \tag{6.4}$$

Given that a process enters state i having been in state s the time before, we imagine that it determines the next state j, to which it will move according to the state i's transition probabilities P_{si1}, P_{si2}, ..., P_{sim}. However, after j has been selected, but before making this transition from state i to state j, the process "holds" for a time τ_{sij} in state i, this holding time τ_{sij} depends not only on the process destination state but also on the previous state and the state it is in now. The holding times τ_{sij} are positive integer random variable each governed by a probability mass function $h_{sij}(\cdot)$.

$$h_{sij}(n) = P(\tau_{sij} = n) \qquad n = 1,2,\ldots \quad s = 1,\ldots,m \quad i = 1,\ldots,m \quad j = 1,\ldots,m \; . \tag{6.5}$$

We assume that the means $\bar{\tau}_{sij}$ of all holding time distributions are finite, and that all holding times are at least one time unit in length,

$$h_{sij}(0) = 0 \ . \tag{6.6}$$

We must specify m^3 holding time mass functions, in addition to the transition probabilities to describe the second order discrete time semi-Markov process.

Holding Times and Waiting Times

It is useful to develop this additional notation. We use $\leq h_{sij}(n)$ for the cumulative probability distribution of τ_{sij}

$$\leq h_{sij}(n) = \sum_{k=0}^{n} h_{sij}(k) = P(\tau_{sij} \leq n) \tag{6.7}$$

and $>h_{sij}(n)$ for the complementary cumulative probability distribution of τ_{sij}.

$$>h_{sij}(n) = \sum_{k=n+1}^{\infty} h_{sij}(k) = P(\tau_{sij} > n) \ . \tag{6.8}$$

Call τ_{si} the waiting time in state i given that the process was in state s the time before, and $w_{si}(\cdot)$ the probability mass function of this random variable. Thus, a waiting time is merely a holding time that is unconditional on the destination state. Then

$$w_{si}(k) = \sum_{j=1}^{m} P_{sij} \, h_{sij}(k) \qquad s = 1,2,\ldots,m \tag{6.9}$$

$$= P(\tau_{si} = k) \ .$$

The mean waiting time $\bar{\tau}_{si}$ is related to the mean holding time $\bar{\tau}_{sij}$ by

$$E(\tau_{si}) = \bar{\tau}_{si} = \sum_{j=1}^{m} P_{sij}\ \bar{\tau}_{sij} \qquad s = 1,2,\ldots,m\ . \tag{6.10}$$

The second moments are related in the same way:

$$E(\tau_{si}^2) = \overline{\tau_{si}^2} = \sum_{j=1}^{m} P_{sij}\ \overline{\tau_{sij}^2} \qquad s = 1,\ldots,m\ . \tag{6.11}$$

Then we can write the variance of the waiting time

$$V(\tau_{si}) = \overline{\tau_{si}^2} - (\bar{\tau}_{si})^2 \qquad s = 1,\ldots,m\ . \tag{6.12}$$

Also the cumulative and complementary cumulative probability distributions for the waiting times are

$$\overset{\leq}{w}_{si}(n) = P(\tau_{si} \leq n) = \sum_{k=0}^{n} w_{si}(k)\ ,\ s = 1,\ldots,m \tag{6.13}$$

$$= \sum_{k=0}^{n} \sum_{j=1}^{m} P_{sij}\ h_{sij}(k) \qquad ,\ s = 1,\ldots,m$$

$$= \sum_{j=1}^{m} P_{sij}\ \overset{\leq}{h}_{sij}(n) \qquad ,\ s = 1,\ldots,m$$

and

$$\overset{>}{w}_{si}(n) = P(\tau_{si} > n) = \sum_{k=n+1}^{\infty} w_{si}(k)\ ,\ s = 1,\ldots,m \tag{6.14}$$

$$= \sum_{k=n+1}^{\infty} \sum_{j=1}^{m} P_{sij}\ h_{sij}(k) \qquad ,\ s = 1,\ldots,m$$

$$= \sum_{j=1}^{m} P_{sij}\ \overset{>}{h}_{sij}(n) \qquad ,\ s = 1,\ldots,m\ .$$

Example 6-1

Consider a second order discrete time semi-Markov process described by the following two matrices:

$$\text{Transition Probability Matrix } P = \begin{bmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \\ 0.7 & 0.3 \\ 0.5 & 0.5 \end{bmatrix}$$

$$\text{Holding Time Mass Function Matrix } H(k) = \begin{bmatrix} h_{111}(k)=\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^{k-1} & h_{112}(k)=\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^{k-1} \\ h_{121}(k)=\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^{k-1} & h_{122}(k)=\left(\frac{1}{12}\right)\left(\frac{11}{12}\right)^{k-1} \\ h_{211}(k)=\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^{k-1} & h_{212}(k)=\left(\frac{1}{7}\right)\left(\frac{6}{7}\right)^{k-1} \\ h_{221}(k)=\left(\frac{1}{5}\right)\left(\frac{4}{5}\right)^{k-1} & h_{222}(k)=\left(\frac{1}{8}\right)\left(\frac{7}{8}\right)^{k-1} \end{bmatrix} \quad k=1,2,\ldots .$$

The holding time distributions are all geometric distributions with different parameters. Then the moments for holding times are:

$\bar{\tau}_{111} = 3$	$\bar{\tau^2}_{111} = 15$	$V(\tau_{111}) = 6$
$\bar{\tau}_{112} = 6$	$\bar{\tau^2}_{112} = 66$	$V(\tau_{112}) = 30$
$\bar{\tau}_{121} = 4$	$\bar{\tau^2}_{121} = 28$	$V(\tau_{121}) = 12$
$\bar{\tau}_{122} = 12$	$\bar{\tau^2}_{122} = 276$	$V(\tau_{122}) = 132$
$\bar{\tau}_{211} = 2$	$\bar{\tau^2}_{211} = 6$	$V(\tau_{211}) = 2$
$\bar{\tau}_{212} = 7$	$\bar{\tau^2}_{212} = 91$	$V(\tau_{212}) = 42$
$\bar{\tau}_{221} = 5$	$\bar{\tau^2}_{221} = 45$	$V(\tau_{221}) = 20$
$\bar{\tau}_{222} = 8$	$\bar{\tau^2}_{222} = 120$	$V(\tau_{222}) = 56$

for example $\bar{\tau}_{121} = 4$ means that given the process was in state 1 the time before, it will stay (on the average) in state 2 for four time units before preceding to state 1 in its coming transition.

The cumulative and complementary cumulative distributions of the holding time can be written in this matrix form:

$$\leq_{H(n)} = \begin{bmatrix} 1 - \left(\frac{1}{3}\right)^n & 1 - \left(\frac{1}{6}\right)^n \\ 1 - \left(\frac{1}{4}\right)^n & 1 - \left(\frac{1}{12}\right)^n \\ 1 - \left(\frac{1}{2}\right)^n & 1 - \left(\frac{1}{7}\right)^n \\ 1 - \left(\frac{1}{5}\right)^n & 1 - \left(\frac{1}{8}\right)^n \end{bmatrix}, \quad n = 0,1,2,\ldots$$

$$>_{H(n)} = \begin{bmatrix} \left(\frac{2}{3}\right)^n & \left(\frac{5}{6}\right)^n \\ \left(\frac{3}{4}\right)^n & \left(\frac{11}{12}\right)^n \\ \left(\frac{1}{2}\right)^n & \left(\frac{6}{7}\right)^n \\ \left(\frac{4}{5}\right)^n & \left(\frac{7}{8}\right)^n \end{bmatrix}, \quad n = 0,1,2,\ldots .$$

These results show, for example that $1 - \left(\frac{1}{12}\right)^n$ is the probability that the process will stay n or fewer time periods in state 1 given that the process was in state 1 the time before, in state 1 now and will be in state 2 in its coming transition.

To find the waiting time statistics one substitutes in Equation (6.10) to get:

$$\bar{\tau}_{11} = (0.2)(3) + (0.8)(6) = 5.4$$

$$\bar{\tau}_{12} = (0.4)(4) + (0.6)(12) = 8.8$$

$$\overline{\tau}_{21} = (0.7)(2) + (0.3)(7) = 3.5$$

$$\overline{\tau}_{22} = (0.5)(5) + (0.5)(8) = 6.5 .$$

This means that $\overline{\tau}_{12} = 8.8$ is the time the process will stay (on the average) in state 2 now given that it was in state 1 last time.

Substituting in Equation (6.11) one gets the second moments:

$$\overline{\tau^2_{11}} = (0.2)(15) + (0.8)(66) = 55.8$$

$$\overline{\tau^2_{12}} = (0.4)(28) + (0.6)(276) = 176.8$$

$$\overline{\tau^2_{21}} = (0.7)(6) + (0.3)(91) = 31.5$$

$$\overline{\tau^2_{22}} = (0.5)(45) + (0.5)(120) = 82.5 .$$

Then one can compute the variances,

$$V(\tau_{11}) = 26.64$$

$$V(\tau_{12}) = 99.36$$

$$V(\tau_{21}) = 19.25$$

$$V(\tau_{22}) = 40.25 .$$

The distribution of waiting can be found by using Equation (6.9),

$$w_{11}(k) = (0.2)(1/3)(2/3)^{k-1} + (0.8)(1/6)(5/6)^{k-1} \qquad k = 1,2,\ldots;$$

$$w_{12}(k) = (0.4)(1/4)(3/4)^{k-1} + (0.6)(1/12)(11/12)^{k-1} \qquad k = 1,2,\ldots;$$

$$w_{21}(k) = (0.7)(1/2)(1/2)^{k-1} + (0.3)(1/7)(6.7)^{k-1} \qquad k = 1,2,\ldots;$$

$$w_{22}(k) = (0.5)(1/5)(4/5)^{k-1} + (0.5)(1/8)(7/8)^{k-1} \qquad k = 1,2,\ldots .$$

Higher Order Continuous-Time Semi-Markov Chains

In the continuous-time semi-Markov process the transitions of the system can occur after any positive, not necessarily integral, time spent in a state. The analysis for second order continuous-time

semi-Markov process is presented in this section. For orders higher than two, the analysis follows in a similar manner.

The second order continuous-time semi-Markov process is just like the second order discrete-time semi-Markov process except for the domain of the holding times. The states occupied on successive transitions are still governed by the transition probabilities p_{sij} of a second order Markov Chain. However, now the time τ_{sij} that the system will hold in state i before making a transition to state j given that the system was in state s last time is a random variable that can take on any positive value, and not necessarily an integral value. One must now use a probability density function to describe the nature of this random variable.

Holding Times

We will adopt the same notation used in the previous section. Therefore, let $h_{sij}(\cdot)$ be the holding time density function for the holding time τ_{sij}. Complete specification of the holding time behavior of the process requires an m by m by m matrix $H(\cdot)$ with elements $h_{sij}(\cdot)$. The kth moment $E(\tau^k_{sij}) = \overline{\tau^k}_{sij}$ of a holding time τ_{sij} is defined by

$$E(\tau^k_{sij}) = \overline{\tau^k}_{sij} = \int_0^\infty \tau^k h_{sij}(\tau)d\tau \ . \qquad (6.15)$$

One can represent the probabilistic nature of a holding time by its cumulative probability distribution ${}^{\leq}h_{sij}(\cdot)$ and complementary cumulative probability distribution ${}^{>}h_{sij}(\cdot)$ defined by

$${}^{\leq}h_{sij}(t) = P(\tau_{sij} \leq t) = \int_0^t h_{sij}(\tau)d\tau \qquad (6.16)$$

and

$$>h_{sij}(t) = P(\tau_{sij} > t) = \int_t^\infty h_{sij}(\tau)d\tau = 1 - {}^{\leq}h_{sij}(t) \ . \qquad (6.17)$$

Waiting Times

Define τ_{si} the waiting time in state i given that the process was in state s the time before, just as we defined it for the discrete-time case. Since τ_{si} is a continuous random variable like τ_{sij}, it must also be described by a probability density function, for which we use the symbol $w_{si}(\cdot)$. The waiting time density function is related to the holding time density functions by

$$w_{si}(\tau) = \sum_{j=1}^{m} p_{sij}\, h_{sij}(\tau) \ , \qquad (6.18)$$

where we have simply weighted the holding time density functions by the probabilities that they will occur.

The waiting time τ_{si} has a kth moment $E(\tau_{si}^k) = \overline{\tau^k_{si}}$ defined by

$$E(\tau_{si}^k) = \overline{\tau^k_{si}} = \int_0^\infty \tau^k\, w_{si}(\tau)d\tau \ . \qquad (6.19)$$

By using Equations (6.15) and (6.18) one can express the waiting time moments in terms of the holding time moments,

$$\begin{aligned} E(\tau_{si}^k) = \overline{\tau^k_{si}} &= \int_0^\infty \tau^k \left(\sum_{j=1}^{m} p_{sij}\, h_{sij}(\tau)\right)d\tau \\ &= \sum_{j=1}^{m} p_{sij} \int_0^\infty \tau^k\, h_{sij}(\tau)d\tau \\ &= \sum_{j=1}^{m} p_{sij}\, \overline{\tau^k_{sij}} \ . \end{aligned} \qquad (6.20)$$

Again, one usually deals with the mean $\overline{\tau}_{si}$, second moment $\overline{\tau^2_{si}}$ and variance $V(\tau_{si})$ of the waiting times.

The cumulative probability distribution ${}^{\leq}w_{si}(\cdot)$ and the complementary cumulative probability distribution ${}^{>}w_{si}(\cdot)$ provide an alternate representation of the probabilistic nature of waiting times. They are given by

$$
\begin{aligned}
{}^{\leq}w_{si}(t) = P(\tau_{si} \leq t) &= \int_0^t w_{si}(\tau)d\tau = \int_0^t \Big(\sum_{j=1}^{m} p_{sij}\, h_{sij}(\tau) \Big) d\tau \\
&= \sum_{j=1}^{m} p_{sij}\, {}^{\leq}h_{sij}(t) \qquad (6.21)
\end{aligned}
$$

and

$$
\begin{aligned}
{}^{>}w_{si}(t) = P(\tau_{si} > t) &= \int_t^{\infty} w_{si}(\tau)d\tau = \int_t^{\infty} \Big(\sum p_{sij}\, h_{sij}(\tau) \Big) d\tau \\
&= \sum_{j=1}^{m} p_{sij}\, {}^{>}h_{sij}(t) \quad . \qquad (6.22)
\end{aligned}
$$

Example 6-2

Considering the same transition probability matrix used in Example 6-1, and specifying the holding time density function matrix as:

$$
H(\tau) = \begin{bmatrix} 4e^{-4\tau} & 2e^{-2\tau} \\ 3e^{-3\tau} & e^{-\tau} \\ 2e^{-2\tau} & 3e^{-3\tau} \\ e^{-\tau} & 2e^{-2\tau} \end{bmatrix}, \quad \tau \geq 0 \, .
$$

The holding time distributions are exponential distributions. The moments of these exponentially distributed holding times are

$$\bar{\tau}_{111} = \frac{1}{4}, \quad \overline{\tau^2}_{111} = \frac{1}{8}, \quad V(\tau_{111}) = \frac{1}{16} \qquad \bar{\tau}_{112} = \frac{1}{2}, \quad \overline{\tau^2}_{112} = \frac{1}{2}, \quad V(\tau_{112}) = \frac{1}{4}$$

$$\bar{\tau}_{121} = \frac{1}{3}, \quad \overline{\tau^2}_{121} = \frac{2}{9}, \quad V(\tau_{121}) = \frac{1}{9} \qquad \bar{\tau}_{122} = 1, \quad \overline{\tau^2}_{122} = 2, \quad V(\tau_{122}) = 1$$

$$\bar{\tau}_{211} = \frac{1}{2}, \quad \overline{\tau^2}_{211} = \frac{1}{2}, \quad V(\tau_{211}) = \frac{1}{4} \qquad \bar{\tau}_{212} = \frac{1}{3}, \quad \overline{\tau^2}_{212} = \frac{2}{9}, \quad V(\tau_{212}) = \frac{1}{9}$$

$$\bar{\tau}_{221} = 1, \quad \overline{\tau^2}_{221} = 2, \quad V(\tau_{221}) = 1 \qquad \bar{\tau}_{222} = \frac{1}{2}, \quad \overline{\tau^2}_{222} = \frac{1}{2}, \quad V(\tau_{222}) = \frac{1}{4}.$$

One can construct the matrices of cumulative holding time probabilities,

$$\leq_{H(t)} = \begin{bmatrix} 1-e^{-4t} & 1-e^{-2t} \\ 1-e^{-3t} & 1-e^{-t} \\ 1-e^{-2t} & 1-e^{-3t} \\ 1-e^{-t} & 1-e^{-2t} \end{bmatrix}, \qquad >_{H(t)} = \begin{bmatrix} e^{-4t} & e^{-2t} \\ e^{-3t} & e^{-t} \\ e^{-2t} & e^{-3t} \\ e^{-t} & e^{-2t} \end{bmatrix}.$$

The waiting time density functions for the example follow from Equation (6.18).

$$w_{11}(\tau) = 0.8e^{-4\tau} + 1.6e^{-2\tau}$$

$$w_{12}(\tau) = 1.2e^{-3\tau} + 0.6e^{-\tau}$$

$$w_{21}(\tau) = 1.4e^{-2\tau} + 0.9e^{-3\tau}$$

$$w_{22}(\tau) = 0.5e^{-\tau} + 1.0e^{-2\tau}.$$

Equation (6.20) allows us to write the waiting time moments

$$\bar{\tau}_{11} = (0.2)(1/4) + (0.8)(1/2) = 0.450$$

$$\bar{\tau}_{12} = (0.4)(1/3) + (0.6)(1) = 0.733$$

$$\bar{\tau}_{21} = (0.7)(1/2) + (0.3)(1/3) = 0.45$$

$$\bar{\tau}_{22} = (0.5)(1) + (0.5)(1/2) = 0.75$$

$$\overline{\tau_{11}^2} = (0.2)(1/8) + (0.8)(1/2) = 0.425$$

$$\overline{\tau_{12}^2} = (0.4)(2/9) + (0.6)(2) = 1.289$$

$$\overline{\tau_{21}^2} = (0.7)(1/2) + (0.3)(2/9) = 0.4167$$

$$\overline{\tau_{22}^2} = (0.5)(2) + (0.5)(1/2) = 1.25 \ .$$

Then the variance of the waiting time is given by

$$V(\tau_{11}) = 0.222$$

$$V(\tau_{12}) = 0.752$$

$$V(\tau_{21}) = 0.213$$

$$V(\tau_{22}) = 0.687 \ .$$

CHAPTER VII

CONCLUSIONS AND RECOMMENDATIONS

This research has been directed towards analyzing higher order Markov and semi-Markov chains. The study revealed that the analysis of such chains using n-dimensional matrices was computationally superior to the existing method found in the literature.

The study of n-dimensional matrices showed that these matrices had different properties as opposed to two-dimensional matrices, mainly:

1) The associative law of multiplication does not hold.

2) For an n-dimensional matrix A and the identity matrix I the following is true:

   a) $A^k = AA^{k-1} \neq A^{k-1} A$ for $k = 2, 3, \ldots$

   b) $AI = A \neq IA$

   c) $A^{-1} A = I \neq AA^{-1}$

It was found that as in the case of the first order Markov chains, steady state probabilities for higher order Markov chains do not depend upon the present and past states of the process. An algorithm was developed to compute steady state probabilities.

The study of the absorption properties of higher order Markov chains revealed that the following quantities can be easily calculated once the inverse of the matrix Q^*_{ii} is obtained:

1) The expected number of steps before the process is absorbed.

2) The expected number of times the process is in a given non-absorbing state.

3) The probability of absorption by any given absorbing state.

The study of higher order discrete time semi-Markov processes revealed that the generality it provides in modeling does not cause unusual computational problems. On the other hand, higher order continuous-time semi-Markov processes do exact a price in computation.

A possible area of further study would be to extend the analysis of Higher Order Semi-Markov chains. To derive equations that would compute interval transition probabilities, entrance and destination probabilities for higher order chains analogous to first order chains.

Another area of possible investigation would be to find more applications and uses for n-dimensional matrices in fields other than higher order Markov chains.

It was found that by permuting the rows and columns of the sparse matrix Q^{*}_{ii} a particular pattern arose which might make the matrix easier to invert. Consequently, another area of possible research is to develop an algorithm for inverting this particular type of sparse matrix.

As a final recommendation it is suggested that the techniques developed in this dissertation be implemented in computer codes so that they may be more conveniently applied to the solution of actual problems.

SELECTED BIBLIOGRAPHY

(1) Anderson, T. W., and L. O. Goodman. "Statistical Inference About Markov Chains." The Annals of Mathematical Statistics, 28 (1957), pp. 89-110.

(2) Bailey, N. T. J. The Elements of Stochastic Processes with Applications to the Natural Sciences. New York: John Wiley and Sons, 1964.

(3) Bartholomew, D. J. Stochastic Models for Social Processes. New York: John Wiley and Sons, 1967.

(4) Bewley, L. V. Tensor Analysis of Electric Circuits and Machines. New York: Ronald Press Company, 1961.

(5) Bharucha-Reid, A. T. Elements of the Theory of Markov Process and Their Applications. New York: McGraw-Hill, 1960.

(6) Bhat, U. N. Elements of Applied Stochastic Processes. New York: John Wiley and Sons, 1972.

(7) Billingsley, P. "Statistical Methods in Markov Chains." The Annals of Mathematical Statistics, 32 (1961), pp. 12-40.

(8) Breiman, L. Probability and Stochastic Processes with a View Toward Applications. Boston: Houghton Mifflin Co., 1969.

(9) Chung, K. L. Markov Chains. New York: Springer-Verlag, 1967.

(10) Clarke, A. B., and R. L. Disney. Probability and Random Processes for Engineers and Scientists. New York: John Wiley and Sons, 1970.

(11) Cox, D. R., and H. D. Miller. The Theory of Stochastic Processes. New York: John Wiley and Sons, 1965.

(12) Doob, J. L. Stochastic Process. New York: John Wiley and Sons, 1963.

(13) Feller, W. An Introduction to Probability Theory and Its Applications. Volume 1, 3rd ed. New York: John Wiley and Sons, 1968.

(14) Fox, L. An Introduction to Numerical Linear Algebra. London: Oxford University Press, 1964.

(15) Franklin, J. N. Matrix Theory. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

(16) Ganesan, S. G. "Analysis of Higher Order Markov Chains." (Unpub. Ph.D. Dissertation, Oklahoma State University, May, 1970.)

(17) Gantmacher, F. R. Application of the Theory of Matrices. New York: Interscience, 1959.

(18) Good, I. J. "The Likelihood Ratio Test for Markov Chains." Biometrika, 42 (1955), pp. 531-533.

(19) Herniter, J. D., and R. A. Howard. "Stochastic Marketing Models." Progress in Operations Research, Vol. 2. New York: John Wiley and Sons, 1964.

(20) Hillier, F. S., and G. L. Lieberman. Introduction to Operations Research. San Francisco: Holden-Day, 1967.

(21) Howard, R. A. Dynamic Programming and Markov Processes. Cambridge: M.I.T. Press, 1960.

(22) Howard, R. A. Dynamic Probabilistic Systems, Vol. I. New York: John Wiley and Sons, 1971.

(23) Howard, R. A. Dynamic Probabilistic Systems, Vol. II. New York: John Wiley and Sons, 1971.

(24) Karlin, S. A First Course in Stochastic Processes. New York: Academic Press, 1966.

(25) Kaufmann, A., and R. Faure. Introduction to Operations Research. New York: Academic Press, 1968.

(26) Kemeny, J. G., and J. L. Snell. Finite Markov Chains. Princeton: VanNostrand, 1959.

(27) Kemeny, J. G., H. Mirkil, J. L. Snell, and G. L. Thompson. Finite Mathematical Structures. Englewood Cliffs, N. J.: Prentice-Hall, 1959.

(28) Kempthorne, O., and J. L. Folks. Probability, Statistics and Data Analysis. Ames, Iowa: Iowa State University Press, 1971.

(29) Kron, G. A Short Course in Tensor Analysis for Electrical Engineers. New York: John Wiley and Sons, 1942.

(30) Kuehn, A. A. "Consumer Brand Choice as a Learning Process." Journal of Advertising Research, 2 (December, 1962), pp. 10-17.

(31) Lee, T. C., G. G. Judge, and A. Zellner. Estimating the Parameters of Markov Probability Model from Aggregate Time Series Data. Amsterdam: North Holland Publishing Co., 1970.

(32) Lee, T. C., G. G. Judge, and T. Takayama. "On Estimating Transition Probabilities of a Markov Process." Journal of Farm Economics, Vol. 47 (1965), pp. 742-762.

(33) Matsushita, Y. "Sparse Matrix Inversion on Illiac IV." Doc. # Center for Advanced Computation, University of Illinois (June, 1968).

(34) Miller, G. A. "Finite Markov Processes in Psychology." Psychometrika, 17 (1952), pp. 149-167.

(35) Nathan, A., and R. K. Even. "The Inversion of Sparse Matrices by a Strategy Derived from Their Graphs." Computer Journal, 10 (1967), pp. 190-194.

(36) Newman, M. "Matrix Computation." A Survey of Numerical Analysis. New York: McGraw-Hill, 1962, pp. 222-253.

(37) Parzen, E. Stochastic Processes. San Francisco: Holden-Day, 1962.

(38) Prabhu, N. U. Stochastic Processes. New York: Macmillan, 1965.

(39) Sikorski, R. Advanced Calculus-Function of Several Variables. Warsaw: Polish Scientific Publishers, 1969.

(40) Telser, L. G. "Least Square Estimates of Transition Probabilities." Measure of Economics. Stanford: Stanford University Press, 1963, pp. 270-292.

(41) Tewarson, R. P. "Computation with Sparse Matrices." SIAM Rev-12 (1970), pp. 527-543.

(42) Turner, J. C. Modern Applied Mathematics, Probability-Statistics-Operational Research. London: The English University Press, 1970.

(43) Walsh, J. "Direct and Indirect Methods." Large Sparse Sets of Linear Equations. [Reid (1970A)], pp. 41-56.

(44) Whittle, P. "Some Distribution and Moment Formulae for the Markov Chain." Journal of Royal Statistical Society, B17 (1955), pp. 235-242.

VITA

Hatem Ahmed Elayat

Candidate for the Degree of

Doctor of Philosophy

Thesis: ANALYSIS OF HIGHER ORDER MARKOV AND SEMI-MARKOV CHAINS

Major Field: Industrial Engineering and Management

Biographical:

Personal Data: Born in Cairo, Egypt, March 25, 1944, the son of Mr. and Mrs. Ahmed Elayat.

Education: Graduated from El-Nasr High School, Maadi, Cairo, Egypt, in May, 1960; received the Bachelor of Science degree from Ain Shams University, Cairo, Egypt, in May, 1965, majoring in Mechanical Engineering; received the Master of Science degree from Oklahoma State University, Stillwater, Oklahoma, in May, 1970, with a major in Industrial Engineering and Management; completed the requirements for the Doctor of Philosophy degree in July, 1973, with a major in Industrial Engineering and Management.

Professional Experience: Instructor, Ain Shams University, Cairo, Egypt, September, 1965, to July, 1968; Graduate Teaching Assistant, Department of Industrial Engineering and Management, Oklahoma State University, January, 1970, to May, 1973.